DOCUMENT RESUME

ED 032 794

EM 007 484

By-Brooks, Lloyd O.
Shaping Faster Question Answering.
American Institutes for Research, Palo Alto, Calif.
Spons Agency-Office of Naval Research, Washington, D.C. Personnel and Training Branch.
Report No-TR-10
Pub Date Jun 65
Note-77p.
EDRS Price MF-$0.50 HC-$3.95
Descriptors-*Achievement Tests, Constructed Response, Instruction, Programed Instruction, *Reinforcement, Rewards, Scoring, *Teaching Machines, *Testing, *Timed Tests

To test a hypothesis that question answering speed and accuracy can be increased by an automated shaping procedure, a film, "The Analysis of Behavior," was presented individually by a teaching machine during twice-per-week sessions to one high school student and 12 junior college students. Six of the students were informed of monetary rewards for increasing speed as well as accuracy in each of their responses to questions on seven sets of the film while the remaining seven students were informed of monetary rewards for response accuracy only. Student responses to multiple presentations of each set were scored for accuracy (the control condition) and, half of the time, for speed as well (the experimental condition) by a machine providing immediate reinforcement. When frames of the experimental condition responses were compared with frames of the control condition, it was found that scoring and reinforcing for speed were statistically reliable in increasing both speed and accuracy; a statistically significant relationship exists between the number of reinforcements and question answering accuracy and speed. In addition, results showed that instructions could increase performance in periods of infrequent reinforcement. (SP)

ED0 32794

AIR-C28-6/65-TR

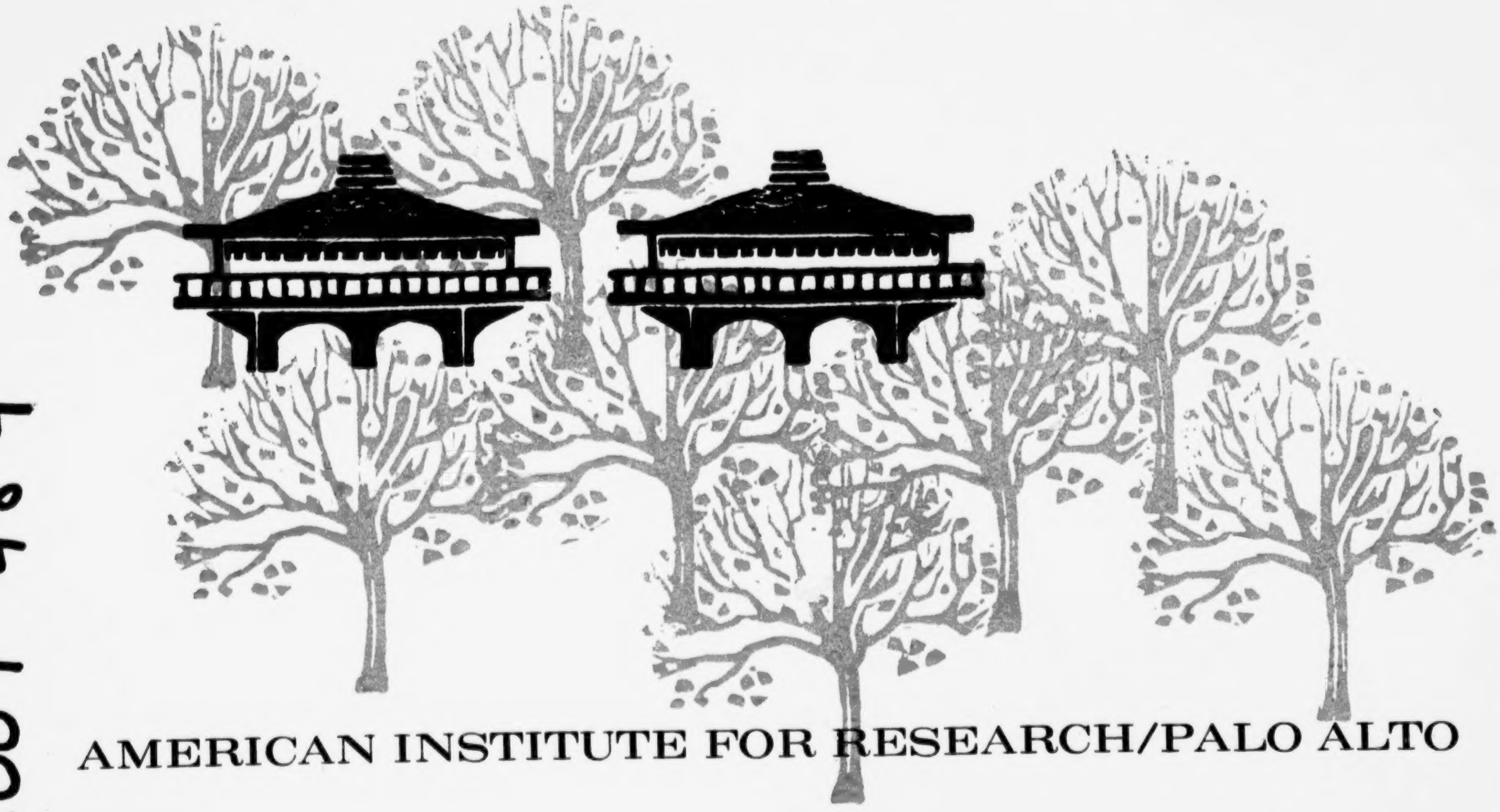
SHAPING FASTER QUESTION ANSWERING

Lloyd O. Brooks

June 1965

Office of Naval Research
Contract No. Nonr-3077(oo)

Technical Report No. 10



AMERICAN INSTITUTE FOR RESEARCH/PALO ALTO

EM007 484

AMERICAN INSTITUTE FOR RESEARCH

mailing address: Station A, P.O. Box 11487, Palo Alto, Calif.
Telephone 321-2130 (Area Code 415)
office location: 1791 Arastradero Road, Palo Alto, California

SHAPING FASTER QUESTION ANSWERING

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Lloyd O. Brooks

AIR-C28-6/65-TR10

Technical Report No. 10
Contract No. Nonr-3077(00)
Leslie J. Briggs, Principal Investigator

June 1965

Personnel and Training Branch
Office of Naval Research
Department of the Navy

SHAPING FASTER QUESTION ANSWERING

ABSTRACT

In laboratory experiments conducted with a dozen junior college students, who participated for one month of twice-per-week sessions at a teaching machine, it was demonstrated possible to increase the rate of answering the questions of early sets of *The Analysis of Behavior* (by J.G. Holland and B.F. Skinner, 1961). The change in rate was brought about by an automated shaping procedure.

Ss responded to a filmed copy of the published version of the program by pressing a key corresponding to the first letter of their answer word, or in case of an answer of more than one word, with the first letter of the last word of their answer. These instructions made it possible to reduce all frames to one-answer questions and to provide automatic machine scoring -- at the expense of fewer than 5 per cent trivial or giveaway questions.

Circuits were devised which made it possible to score, automatically, answers given correctly before some preset value of time had elapsed. Preset values differed from question to question and were estimates of average times based on data from other groups of *Ss*. Scores appeared before *S* in counters, and a change in score in one counter was always accompanied by the flash of a light. In addition to a per-session honorarium, *S* received one cent for each correct answer and another cent for each correct answer given before the preset time for a particular question had elapsed, provided he continued in the study until its completion.

Each point for speed was accompanied by a reduction in preset times for all questions yet to be answered. Each failure to obtain a point, whether owing to error or to slow answering, typically reset all times to the initial values. It was possible for *S* to obtain points for speed only during either the first or last half of each lesson, but the time when he could obtain points was not indicated to him by instructions or by stimulus change. Treatment and control conditions were counterbalanced.

A cumulative recorder and printing counter provided the major results, both of *S*'s performance and as monitors of the adequacy of critical equipment functions. In addition, *Ss* were interviewed; and voluntary comments were encouraged throughout the study.

The time spent completing the last ten questions in the control condition (points *not* given following correct answers given quickly) was compared with time spent on the last ten questions in the experimental condition (points *given* for correct answers given quickly), after correcting for any inherent inequality in the time required for the two ten-question units of material.

The major finding was that scoring for speed was statistically reliable in increasing speed *on the first presentation* of a lesson. On subsequent presentations of the same lesson, *Ss* seemed to answer questions faster

part of the time but even slower than previously some of the time. Performance suggesting such disruption seemed to continue for as many as four repetitions of a lesson, with gradual reduction of disruption -- when this many repetitions were considered for two lessons.

On the basis of interview data, volunteered comments, cumulative records, and the kind of program questions typically associated with performance disruption, it would seem that Ss came to answer correctly questions, when repeated, increasingly faster by learning the prompts and by remembering answers by associating them with key words in the question. As a result, when items were encountered from which the prompts had vanished, especially if the answer required a consideration of the entire statement, S was forced to deal with the substance of a question, now somewhat out of context as a result of his attending to prompts rather than subject matter in the development leading up to it. An implication of this finding is that programs should not be repeated as a means of review. Variant forms of a program employing different kinds of weaker prompts would be superior.

Of the more than four thousand correct answers recorded, 2,236 were given for the questions of the control condition and 2,232 for questions of the treatment condition. When subtotals of correct answers in each condition were considered, sometimes more correct answers were given in one condition than another. Such variation in differences between treatment and control conditions in the number of correct answers throughout the study suggested that the scoring of Ss' answers for both speed and correctness brought about increases in the number of correct answers whenever it brought about increases in Ss' speed of answering.

During the course of the study, half of the Ss were told that the points in the #2 counter were based on speed of correctly answering questions and on improvement in this speed. The other 6 Ss were merely told what the points were worth. Although some Non-Instructed Ss seemed responsive to the experimental treatment, and there was little relationship between performance and what Ss reported they regarded as the contingency between bonus points and performance, no statistically reliable difference between Ss' treatment-and control-condition performances was found in the results of the Non-Instructed Group, when data were analyzed for this group separately.

It is difficult to estimate the gain possible in improvement if the treatment were prolonged but applied only to the first presentation of lessons. Many Ss reported making an effort to improve speed throughout each lesson. Consequently, the rate during the control condition may be high due to a failure to discriminate between when the treatment was or was not in effect. The present study should underestimate treatment effectiveness.

While repeating a lesson limits the treatment effect, task difficulty may not be a limitation, as long as S is able, eventually, to respond correctly. The gains may even be greater proportionately for more difficult tasks.

TABLE OF CONTENTS

Page

INTRODUCTION 1

Related Research 3

THE PROBLEM 8

METHOD 9

Design 9

Procedures 10

Equipment 10

Subjects 11

Instructions 11

Machine scoring 12

Sets or lessons 14

Faulty procedures or equipment failures 15

Preset values 16

RESULTS 17

Kinds of Analyses 18

Repetitions of lessons 18

Effect of first instances of experimental treatment on criterion time measures 20

Effectiveness of instructions 21

Alternative corrections for inequality of read-and-answer times associated with the first and second halves of each lesson 22

Number of correct answers 23

Number of bonus points 24

Effects of preset values related to instructions 25

Accommodating individual differences 27

Read-and-answer times at the moment of reinforcement 27

TABLE OF CONTENTS
(continued)

Page

DISCUSSION (Major Results) 30

OTHER RESULTS 35

    Apparent Ineffectiveness of Treatment
        During Lesson Repetition 35
    Interview Data 37
    Response Modes 40

FURTHER RESEARCH 41

SUMMARY 43

APPENDIX 47

REFERENCES 59

TABLES

Page

Table 1 Mean Differences in Read-and-Answer Times Between Control and Experimental Treatments for Instructed and Non-Instructed Subjects 19

Table 2 Difference Scores (Control Conditions Minus Experimental Treatment) for First Scoring of Correct Answers for Speed During a First or Second Half of the Lesson 52

Table 3 Difference Scores (Control Condition Minus Experimental Treatment) for First Scoring of Correct Answers for Speed During a First or Second Half of the Lesson (Results Arranged for Major Comparisons) 53

Table 4 Difference Scores (Control Condition Minus Experimental Treatment) for Second Scoring of Correct Answers for Speed (Third Presentation of Sets 2, 3, 4, 5-1, and 5-2) 54

Table 5 Difference Scores (Control Condition Minus Experimental Treatment) for First Scoring for Speed Added to Corresponding Difference Scores for Second Scoring for Speed 55

Table 6 Alternative Corrections for Inequality of Lesson Halves in Obtaining Difference Scores (Control Condition Minus Experimental Treatment) for First Scoring of Correct Answers for Speed During a First or Second Half of the Lesson 56

TABLES
(continued)

Page

Table 7 Average Read-and-Answer Times for Frames Answered Correctly by S and Followed by #2-counter Points During the First Scoring for Speed 57

Table 8 Average Read-and Answer Times for 20 Lesson Frames, on which Criterion Mean Differences were Later Obtained, when Completed in Composed-Answer Mode During First Presentation 58

FIGURES

Page

Figure 1 The special procedure of presenting Set 7 to S #12 five times. 36

Figure 2 Set 5-1 was presented to S #8 in composed-answer mode followed by two presentations with machine scoring. This Non-Instructed S asked, at the end of this session, "What were the points for?" 39

SHAPING FASTER QUESTION ANSWERING[1]

Lloyd O. Brooks
American Institutes for Research
Palo Alto, California

INTRODUCTION

In a recent state and national survey, sponsored by a USOE Title VII grant and conducted by the Pennsylvania Department of Public Instruction, the following conclusion was based on teacher and administrator responses to attitudinal questionnaire items: "...reading comprehension generally improves more with programed instruction than it does with conventional instruction" (Archer & Sanzotto, 1964, pp. 608-609). If this is true of programed instruction in general, possibly it would be especially true of programed learning accomplished with a teaching machine. Holz and Robinson (1963) suggested that greater control is possible with machine than text (indicated by a higher correlation between errors on a review portion of a program and errors on a postprogram test, and by a higher error rate for the program when presented by machine rather than as a text). The greater control possible with a teaching machine also provides a basis for the improvement of other aspects of reading skill.

The laboratory experimentation described in this report began as an attempt to determine whether reading speed could be increased as a by-product of teaching machine instruction. It was reasoned that machine control could be increased by adding features such as a bonus score

---

[1] The laboratory experimentation described in this report was largely supported by a grant from the American Institutes for Research (RD-28). It is gratefully acknowledged that part of the capability for this kind of experimentation is the result of support for research concerning response modes and feedback variables in programed instruction under Contract Nonr-3077 (00) with the Office of Naval Research, L. J. Briggs, principal investigator. Analyses of results, particularly as these pertain to a comparison of response modes, and the report preparation and distribution were largely supported by this contract of the American Institutes for Research with the Office of Naval Research. Mrs. Beverly Adkins was of substantial help in data reduction and statistical analyses. Mrs. Dorothy Hanson, Director of Foothill College Student Placement, was especially helpful in arranging the scheduling of students.

following those correct answers given quickly. By monitoring errors as well as performance rate it might be possible to increase the student's reading rate safely, short of the point that led to an increase in his errors.

As an approach to reading improvement, several points of departure were evident. First, reading improvement was to come about as a by-product of various other educational activities and not as a special activity carried on at the expense of these.

Second, the frequent opportunity for some kind of overt error afforded by programed materials, present even in low-error-rate programs if reading "comprehension" were impaired, made it possible, as already suggested, to seek a gain in speed while avoiding performance impairment since any increase in overt errors could alert the experimenter immediately to this undesirable consequence of an experimental procedure.

Third, the use of diverse subject matters and techniques of programing should help in generalizing any gains in reading skill from the context in which these were acquired to the various contexts of usefulness. Moreover, since reading speed was to grow out of a shaping process, rather than left merely to grow out of task mastery without special provision or to result from some form of forced pacing, it was possible that speed would be inherently a part of the learning process rather than a matter of special circumstances. Quick answering could be learned along with correct answering, and both aspects of performance gain in generalizability through occurence in a wide variety of situations.

Fourth, the use of explicit procedures and reinforcers could make it possible to establish faster performance according to one schedule of reinforcement and then increase the persistence of this improvement by shifting to a schedule likely to result in increased resistance to extinction (cf. Ferster & Skinner, 1957, in the context of lower-organism research). Relatedly, attention to matters of reinforcement schedules and their effects on emotional behavior could lead to the evolution of procedures which avoided potential performance disruption, such as could result from abrupt changes in certain reinforcement contingencies.

Holz (1964) reported, in using reinforcements consisting of points, which were later worth money, with college student Ss, that modifications of a schedule were made before the result showed continuity with previous findings. It should be noted that a carrying over of procedures from one context to another is more likely to help suggest to the experimenter procedures to try and changes to look for in behavior than to function as simple application and extrapolation.

The major change in formulation which has resulted from second thoughts during the experimental work to be described in this report is that "reading" is likely to be a misleading oversimplification. To the extent that question answering and reading are equated it is likely, among other consequences, that some of the generality of the procedure may be obscured. Certainly it is unlikely that procedures which reduce question-answering time do so only by reducing the reading portion of the time. It may be profitable to anticipate greater generality of the procedures by regarding question-answering as merely an example of complex performance.

Related Research

In an earlier study (Brooks, 1961), a limited attempt was made to manipulate question-answering time by a shaping procedure in an effort to increase the control of question characteristics. These early efforts were directed toward improving student pacing in the belief that students who made more errors often tended to answer some questions too quickly. Occasionally, students who made few errors seemed to answer questions too slowly.

Maccoby and Sheffield (1961) suggested that self-pacing in learning which involved observation of a demonstration and practice was more a matter of learners adjusting their distribution of demonstrations and practice to their abilities than a genuine training procedure. They suggest the likely superiority of controlled-pacing training procedures selected by the trainer rather than the trainee when these are adjusted for learner ability. Maccoby and Sheffield also suggest that the superior learner is more likely to use appropriate self-pacing techniques than one of lower ability.

Recently (Kress & Gropper, 1964) it has been reported that students who work programed lessons at a pace appropriate to their ability level attain higher achievement scores.

The procedures of the present study could as readily be applied to training in appropriate pacing as merely to increasing speed. It could be desirable to train students in appropriate pacing first in order to decrease their error rate so that there later would be more opportunities to reinforce correct responses given quickly.

Numerous reviewers and researchers have dealt with the topic of response mode as an influence on student, or learner, performance (e.g., Alter & Silverman, 1962; Briggs, Goldbeck, Campbell, & Nichols, 1962; Burton & Goldbeck, 1962; Cummings & Goldstein, 1962; Eigen & Margulies, 1963; Goldbeck, 1960; Goldbeck & Briggs, 1962; Goldbeck & Campbell, 1962; Goldbeck, Campbell, & Llewellyn, 1960; Hamilton, 1964; Hershberger, 1963 a, b, 1964, and Hershberger & Terry, 1963; Holland, 1960; Kidd & Micocci, 1964; Krumboltz & Weisman, 1962; Lumsdaine, 1961, McGuire, 1955; McNeil, 1962; Williams, 1963; and Wittrock, 1963). Feedback variables also have been considered as an important influence, often in relation to response modes, and in relation to other variables (Briggs & Hamilton, 1964). The present study relates to both topics.

The form of the response required of Ss during this study made possible machine scoring of their answers to a composed-answer type of program. As used, multiple-answer frames were converted into single-response frames, although it would have been possible with special filming of the material to present the same frame as many times as would be necessary to get a response to all blanks -- a response to a different blank each time.

For some lessons, both the composed-answer form of response (originally intended by the authors of the program) and the machine-scorable form of response were required, with composed answers given on the first presentation and the machine-scorable responses on the second and third presentation of the same lesson.

The feedback used was that which was originally provided by the program authors, namely, a view of the correct answer after S had either written his answer or pressed a lettered button. The machine-scoring mode provided additional feedback in the form of a flash of light when a counter changed as it added to the score which would eventually be of monetary value. This form of feedback was used primarily for its reinforcement possibility, although it can be expected to be more immediate than S's self-scoring. Also, it provided one form of feedback pertaining to speed and another, redundantly, indicating answer correctness (which S could determine for himself as he saw the program answer and which was a necessary condition for answers to be scored as fast). The accumulated total of a counter, however, is not the same kind of feedback as the individual instances of the counter changing, a complication if S sometimes ignores a counter.

Hershberger (1963 a & b, 1964, and Hershberger & Terry, 1963, cited above) has suggested that reading is an important response mode and has considered, in experimental studies, typographical cueing as a means of influencing reading effectiveness. The present study may also be regarded as concerned with reading as a response mode. However, the reservation suggested earlier of "reading" being an oversimplification should apply here as well.

A major difference between reading as a response in Hershberger's work and in the present is that Hershberger attempted to increase the effectiveness of this kind of response by new techniques of programing, whereas in the present study there is an attempt to increase its efficiency without loss in effectiveness, by adding to the control characteristics of the machine rather than to those of the program. Hershberger examined reading as a non-overt response mode through performance on a test after the program. In the present study, overt responding follows closely on likely instances of reading. Criterion-performance consideration is restricted for the moment to study of performances within the program; concern for postprogram performance is merely deferred for later study. (Cf. Joint Committee Report, 1963; Rothkopf, 1962, intrinsic versus extrinsic criteria of program effectiveness.)

Goldiamond has interpreted reading as a form of recognition or monitoring behavior and suggested that Holland (1957, 1958) has contributed methods for dealing with the problems posed by the lack of an explicit response. Goldiamond presents a method for defining silent reading which involves machine presentation of materials. In contrast to machine control, Goldiamond notes that, "It should come as no surprise that during silent reading, S can gloss over difficult sections. The superiority of machine programmed presentation of identical programmed material lies precisely in elimination by the former of such difficulties inherent in subject control over the program" (p. 11). Goldiamond also suggests the possibility of using machine-defined reading as a useful baseline behavior in research. He cites disruptions of speech and reading as early diagnostic indicators of neurological disturbance.

The possibility of using programed instruction as a means of establishing useful background behavior that is likely to be sensitive to the effects of other variables is also not be overlooked. As the materials used for establishing such behavior are used in obtaining various kinds of normative or baseline data (Brooks, 1964) and machine-definition of performance becomes more extensive, it should become increasingly possible to study basic, general principles in the context of complex human behavior.

Automated procedures for altering response requirements on the basis of immediate performance have received recent consideration (e.g., Ferster & Skinner, 1957; Field & Boren, 1963; Findley, 1958; Hodos, 1961; Hodos & Kalman, 1963; Sidman, 1962; and Verhave, 1963). Adjusting schedules have been used to adjust, e.g., the number of responses required of an animal for each reinforcer. If performance suggests a decrement, as by a long pause after reinforcement, fewer responses are then required for a next reinforcer.

Symptoms of excellent performance may have the opposite effect; the ratio of responses per reinforcer may be changed by increasing the response requirement each time a pause does not occur immediately after reinforcement. The adjustment may be only an increase in response requirement and never a downward adjustment on the basis of a performance decrement (Hodos, 1961; Hodos & Kalman, 1963, cited above).

The possibility of using adjusting procedures as a means of accommodating individual differences and moment-to-moment variation in performance in the context of programed instruction deserves consideration.

Crossman (1965) has suggested that dynamic contingencies which provide adjustments within a single experimental session have received limited study and that an adequate vocabulary is not yet established. Crossman suggests that dynamic contingencies fall largely into two categories. One of these has already been considered: adjusting schedules. These are characterized by the change in a response characteristic (pause after reinforcement, long latency, interresponse time, etc.) bringing about some kind of change in the response contingency (number of responses, length of temporal interval in which a response must occur to be reinforced, vigor of a response in pounds of force, etc.).

Adjusting schedules are to be contrasted with titration schedules. In titration schedules, a change in some response characteristic alters some characteristic of a stimulus (intensity, duration, frequency, etc.). Crossman points out that Blough (1958, 1963) has used an adaptation of a technique developed by Bekesy (1947), derived from the psychological method of limits, and has applied it as a titration schedule in obtaining spectral sensitivity thresholds from pigeons (Blough, 1958) and monkeys (Blough, 1963).

Lumsdaine (Lumsdaine & Glaser, 1960) has suggested examples in the context of programed instruction which can be related to the titration category. Lumsdaine discussed the manipulation of physical properties of prompts and cited the work of Israel, then in progress, as relevant. Lumsdaine also noted that Pask had described an application of such prompting (cf. Israel, 1960; Pask, 1959).

In the present study, each reinforcement of a response that fell within the limit of a temporal interval defined by a preset value altered preset values proportional to their initial settings (adjusting schedule). Each reinforcement could conceivably have altered the strength of prompt for those of a given topic by a decrease along some physical dimension, such as light intensity (titration schedule).

Time has been considered as an important dimension of performance for describing behavior (Berliner, Angell, & Shearer, 1964; Kershner, 1964), particularly in relation to other aspects of behavior, such as error rate (e.g., Alter, 1963; Briggs, Campbell, & Brooks, 1964; Brooks, 1961; Brooks, 1964; Jacobs, 1963; Siegel, 1964; Suppes, 1964; and Tate, 1948). Other studies, such as the present, have been concerned with the control or manipulation of the time dimension of performance (e.g., Goldiamond; Kershner, 1964, cited above).

Coulson (1962), and others, have considered computers as useful devices for presenting instructional programs. Computers probably provide the best capability available for conducting experimental studies of influences on the effectiveness of programed instruction. Strollo's work (Strollo, 1964) cited by Clapp, Yens, Shettel, and Mayer (1964), which was discovered after the present study was completed, is probably the most relevant to it of all computer applications.

The present study lies somewhere between the use of a full-scale computer and the use of a simple teaching machine. By developing control circuits which add to the capability of a simple device, the result has been to develop a simple form of highly specialized computer which has been designed especially to handle certain kinds of non-subject-matter branching. Such less-expensive devices can provide for extensive pilot research, as in the present study, at low cost and help define the kinds of problems that only the more complex computer can help solve.

As a final comment on related research and methodological matters, it should be noted that the procedures used in paying S's honoraria, based on performance but contingent upon successful completion of the experiment, were rather effective. Twelve of the thirteen Ss remained in the study until its completion. The one S lost along the way earned unusually few bonus points. This suggests that it would be especially important to accommodate individual differences to the extent that all Ss did accumulate relatively equal bonuses. Otherwise, attrition could become selective and biasing. Staats, Minke, Finley, Wolfe, and Brooks (1964) suggest procedures useful in obtaining long-term participation of young children in laboratory studies.

THE PROBLEM

The problem in the present study was to develop automated procedures which would shape faster question answering as a by-product of study of programed material at a teaching machine. This called for an experimental evaluation under controlled conditions. On the one hand, the attainment of a successful demonstration of an experimental procedure was a basic objective. On the other hand, it seemed that more might be gained by somewhat extensive study of only partially-effective procedures. Although the objective was a definite outcome, arriving at a successful procedure was only slightly more important than discovering conditions which rendered a likely procedure less effective.

METHOD

A filmed, 35mm, copy of the published version of The Analysis of Behavior by J. G. Holland and B. F. Skinner (McGraw-Hill, 1961) was presented (with the publisher's kind permission) individually to junior college students by a teaching machine which automatically scored their answers for correctness and, for half of each lesson, for speed as well. The time spent reading and answering each question by each experimental subject was automatically compared with a value estimated as near average on the basis of student data obtained in a previous study (Briggs, Campbell, & Brooks, 1964).

Design

The main experimental treatment, B, was scoring for speed. The control condition, A, was the absence of this scoring. Conditions for half of the Ss were ABBA, with BAAB for the other Ss during two presentations of the lesson given in immediate succession. On rare occasion lessons were presented several times (ABBABAAB, e.g.). Since the experimental treatment was not equally effective on first and second presentations of the lesson, and since first and second halves of lessons were not equal in the time Ss spent on them when conditions were the same throughout the entire lesson, lesson presentations were considered separately in analyses of results and a correction was made for inequality of lesson halves.

Two approaches to this correction were considered, as will be explained later. One correction was based on composed-answer data from a previous study and such data from some lessons in the present study. Another was derived from data of the present study by comparing the sum of an equal number of criterion measures from Ss in the control (A) and treatment (B) conditions for a first half of the lesson with a sum on the second half for these same Ss with A and B conditions then reversed. The per cent difference in sums could then be applied as a per cent correction in individual criterion measures. This per cent of each S's total was added to the half of his lesson found, using data from all Ss, to require less read-and-answer time.

Criterion values were the times spent reading and answering the last 10 frames of each condition, A or B (control or treatment). Errors on all frames of each treatment condition -- not just the last 10 -- were also considered either directly, or indirectly by considering the number of correct answers.

Procedures

Each S worked individually in a small corner room where he could be observed by E from an adjacent room through a one-way vision window. A masking sound was provided by a tape recorder and speakers outside S's room.

Equipment. Recording and control equipment were located in E's room. Reading-and-answering times in seconds were recorded separately for each frame of material by means of an Elmeg printing counter (Grason-Stadler, E12505A). A partially redundant record was provided by a Gerbrands cumulative recorder (C-3) which gave not only a cumulative record of S's progress, but also marked such special events as S's receiving a score for answering correctly, and for answering correctly and quickly. In addition to displaying gross changes in criterion performances, these records were useful in monitoring the critical functions of control equipment.

Response times in relation to experimental treatment were obtained from printing-counter records for systematic analysis, rather than given merely in the graphic form of cumulative records, since the rate difference in the two conditions (A versus B) tended to be less than 10 per cent. Herrick (1965) comments on the usefulness of cumulative records under such conditions, even in the case of responses of homogeneous typography. "If, for example, E wanted to show that an experimental variable caused a rate increase of 10%, it would be useless to collect the data in the form of a cumulative curve because, even with the best selection of scales, such an increase would be represented by an angular increase less than 2°. Thus, instead of clearly exposing the 10% rate increase, E would be transforming the finding into a visual acuity task for the reader" (p. 60).

S worked facing a DuKane Reditutor viewer, and E was able to see the screen of this viewer from behind and over the shoulder of S as S worked. It was possible to use this viewer with the DuKane Paper Machine (write-in answer unit) attached, or with a full-alphabet keyboard and fully automatic machine scoring. The keyboard device, devised by E, served the main purposes of the present study.

Control circuits, also devised by E, not only made possible the automatic scoring of each keyboard button press for correctness, but also automatically scored correct answers as occurring before or after some preset time had elapsed, a time which typically differed from frame to frame and which was estimated from the data of the prior study.

Scores for correct answers registered in counter #1 in view of S. Scores for correct answers given before some preset time had elapsed registered in counter #2. Counter #3 gave the cumulative total of questions encountered, as a possible aid in S's interpretation of his other scores. Each correct answer was accompanied by the sound of a changing counter (#1) and by a flash of a light mounted on the keyboard.

Late in the study the light was made a function of the change in the #2 counter instead of the #1. During all of the study, each time S gained a point in the #2 counter all preset values were reduced, thereby requiring faster correct answering in order to obtain another bonus point. Failure to obtain a point -- whether because S answered incorrectly, or because he answered correctly but not quickly enough -- reset the preset values to their initial values during most of the study. Late in the study, for Set 7 of the program, this reset feature was eliminated.

Subjects. A total of 17 high school and college students participated in the study, although the main study is based on data of the 12 junior college students who participated in two one-hour sessions per week for one month. Of these, there were 11 males and one female. Three high school female students and one male participated for only one session, the latter providing help merely in testing equipment and procedures. One other junior college student participated for three sessions.

Instructions. All of the college students were told they would receive two dollars for each experimental session of approximately one hour length, and six were told they could earn one cent for each correct answer, plus one additional cent for each correct answer given faster than the average student from whom the norms had been derived (Instructed Group). As they worked they would also have to show improvement in order to continue to receive bonus scores. Seven of the Ss were merely told that each point in either the #1 or #2 counter was worth one cent (Non-Instructed Group). Bonus scores, or points, would only be converted into dollars and cents if the S participated in the study until its completion or until the end of the month, whichever was sooner.

Machine Scoring. S was told that he would sometimes write full answers for frames of programed material. However, in using the keyboard answer unit he was to press the first letter of the last word of his answer (converting numerals into words if necessary). This procedure both made automatic scoring possible and converted all frames into a presentation of only one question. (It is obvious that this will occasionally fail to provide a critical discrimination, as in the context of emit versus elicit, smooth versus striate, or response versus reflex. However, inspection of the portion of the program used in this study suggested that less than 5 per cent of the frames would be converted into trivial or giveaway questions by this procedure.)

Questions could occasionally be answered incorrectly by S's pressing some letter other than the first one of the correct answer term, and yet be scored as correct. This resulted from the fact that it was not possible, in terms of the apparatus available at the time, to connect each of the twenty-six pushbuttons to a separate scoring circuit. Instead, several pushbuttons were connected together, making up seven or eight groups. S could thus be expected to be correct by chance, (by non-systematic influences on performance) about one time in eight. In addition, since all alternatives allowed by the program were also allowed for machine scoring, it was sometimes possible for S to have as much as a 50-50 chance of being correct, although this was unusual. Other features of the circuit could infrequently contribute to an increase in probability of success by machine scoring. Despite these possible complications, statistically significant rank-order correlations between errors on a first presentation and errors on a second and third presentation combined were found for two of the five lessons considered, and rather large correlation-coefficient values for the other three.

For these correlations, the x-scores were the number of frames on which there was at least one error when S wrote his answer; answer tapes were scored by E. The y-scores were the number of instances S failed to receive a point for his answer, as indicated to E by the difference between the number of frames in the lesson and the number of special-event marks in the cumulative record. These were pooled for the second and third presentations of the lesson, for the sake of obtaining greater variance in y-measures than only a second presentation afforded.

For Sets 2, 3, 4, 5-1, and 5-2 all correlation coefficients were of the same sign; two were statistically significant (for Sets 2 and 4), and three not (at the 5% level, one-tailed, all corrected for ties): Set 2, $R_s = .66$, $N = 12$; Set 3, $R_s = .41$, $N = 12$; Set 4, $R_s = .76$, $N = 8$; Set 5-1, $R_s = .47$, $N = 10$; and Set 5-2, $R_s = .69$, $N = 7$. The value of N varies as a result largely of faulty machine score recording.

On the basis of a previous study (Briggs, Campbell, & Brooks, 1964) it was possible to consider scores on a first and second presentation of Sets 4, 5-1, and 5-2 when answers were written on both presentations. Frames were scored by E as incorrect if any part was missed, as was the case of the x-scores above. Respectively, $R_s = .95$, $N = 5$; $R_s = .80$, $N = 5$; and $R_s = .98$, $N = 5$. For Set 5-1 R_s was not statistically significant.

Although scoring in the present study could have been adversely affected by such factors as procedural errors, equipment malfunction, and high chance expectancy, it does not appear to have been affected to an extent which would invalidate it as a scoring method. The correlation is lower, as one would expect, when written-answer scoring by E and machine-scoring modes are compared, especially when it is recognized that any one of several responses to some of the frames when scored by E count as an error for that frame and that machine scoring only is based on a pooling of scores for a second and third presentation.

S's were told they would sometimes intend an answer which was incorrect, but which began with the first letter of the correct answer, and yet receive correct-answer credit for it. Also, sometimes they would receive more than one point in one or both counters for only one answer. Each S was told that as long as the number of points in either counter did not exceed the total number of questions, he would be able to receive one cent for each point in each of the two counters.

A circuit element could be adjusted so as to be biased in S's favor (sensitive to quick changes in current flow); or, omissions of points due S could be correctly credited on the basis of the record available to E if the sensitivity of the circuit element were decreased. Scoring which was generous to S -- less likely to fail to give him credit but more likely to double his credit at times -- would be immediate, less disruptive, and less complex. When E considered scores for the sake of error analyses, he was able to recognize instances of double points and consider them as only one correct answer. These instances seemed rare, probably occurring in less than one per cent of the correct scores.

Sets or Lessons. Set 1 was presented to most of the Ss in the machine-scored mode when it was first presented. Procedures and instrumentation were somewhat provisional. Close scheduling of Ss resulted in incomplete data when an attempt was made to present the lesson in the composed-answer mode first and the machine-scored mode next to some Ss. Most of these Ss had to be interrupted by E and started on the lesson in the machine-scored mode in order to provide data of primary relevance to the objectives of the study within the scheduled hour.

For Sets 2, 3, 4, 5-1, and 5-2 all Ss first wrote their answers, then immediately worked through the same lesson two more times in the machine-scored mode. There was to be complete counterbalancing of conditions for each pair of Ss, provided the assumption of linear effects had been valid.

Sets 6 and 7 were worked only in the machine-scored mode. During Set 6, as was true of all previous lessons during machine scoring, preset time-comparison values were adjusted back to initial settings each time S failed to obtain a bonus score for speed -- whether due to an incorrect answer or to a correct answer not being given soon enough. During Set 7 preset values were adjusted only downward, or smaller, each time S answered correctly in less time than was provided for by these values.

Set 5 was treated as two sets or lessons. It was a longer lesson than most, and originally the first few frames were poorly filmed. Consequently, the first half was refilmed as a separate lesson. When Set 5-1 was presented S wrote his answer the first time and then worked through this portion of Set 5 twice in the machine-scored mode, just as he had done in the case of Sets 2 - 4. On the next session, E advanced the second-half portion to the point where the remaining one half of Set 5 began, and S then wrote his answers on the first presentation of Set 5-2. However, when S next began his two repetitions of the lesson in the machine-scored mode, E started him at an earlier point, on frame #6, at a point just beyond the damaged portion of the film. This made it possible to consider a prolonged treatment condition. However, it also resulted in an unequal number of presentations of portions of the material for this meant that most of 5-1 was repeated a third and fourth time. With linear effects of the experimental treatment this would have been less of a complication, owing to the counterbalancing provided by the experimental design.

Most of the failures to record correct answers in the machine-scoring mode occurred on Set 6. Otherwise, experimental procedures and equipment functions had been greatly improved by the time of this lesson. After one S had completed Set 6 E realized that Ss did not always attend to the counters. The flash of light was an essentially unavoidable stimulus change. Consequently, it would seem a better procedure to correlate the light with the score for speed. The change was made in time for the remaining eleven Ss for Set 6 and all twelve Ss for Set 7.

On Set 7, any decrease in preset comparison values which had resulted from Ss' gaining points for speed was not reversed by his failure to obtain such a point. This would seem to be a form of progressive interval schedule (cf. Hodos, 1961).

The complications of a pacing feature and an irregular cycle starting point, mentioned in the next section, pertain only to the first few lessons.

Faulty procedures or equipment failures. The recording procedures made it possible to record many critical aspects of experimentation. However, the fact that a timing cycle began at random points in the first cycle, introducing imprecise scoring for speed of correct answers, was only noted and corrected after the first few sessions. Since the preset values for most questions were several cycles in length, this was a relatively minor error in circuit design.

At the beginning of the study the seven Ss of the Instructed Group were told that if they answered too quickly, suggesting thereby some guessing rather than careful reading of the material, they could fail to earn a point for speed. When lessons were repeated the third time, one S seemed actually to be reading too fast to receive points for speed because of the pacing feature. This S, incidentally, was the fastest of the group, and usually responded to all frames correctly when lessons were repeated. The slowest S in the study was observed to spend seven times as long as this S on some lessons. Early in the study the inability to accommodate this range of subject differences on multiple presentations was accepted in order to get on with the main purposes of the experiment, and the possibly-unnecessary pacing feature changed so as to place no limit on how fast Ss could answer correctly during the treatment condition and yet obtain a bonus point.

Errors of procedure were infrequent, and none was fatal to the purposes of the study. If, e.g., E failed to start the treatment (a matter since left to the equipment rather than E) at the midpoint of a lesson, this was indicated in the record and the appropriate 10 frames could be used for criterion measures, and first half-second half inequality could be properly allowed for in analyzing the results. The point of change between A and B first came to be recorded by E, then later controlled automatically by his programing circuits (too late for the results of this study, however).

Scoring and timing circuits were usually tested according to all acceptable alternatives and preset-value limits by E. Nevertheless, with extreme changes in temperature and humidity early and late in the day there were occasional instances of equipment malfunctioning.

During the study the special-event pen of the cumulative recorder, which was activated by correct-answer scoring, occasionally failed. For legible recording this pen had to be activated for only a few milliseconds; otherwise the paper travel during its excursion would be sufficient for a double line to be formed by its advance and return, with a messy ink fill-in. Apparently, relay contact wear during the near-full month of use had been sufficient to shorten the control pulse excessively. Consequently, there were sessions for which data were not obtained concerning how many correct answers were given in the A versus B condition, particularly for Set 6.

Preset values. Read-and-answer times were preprogramed for each question, or frame. For some sets these values were all proportionately somewhat less than the values estimated as average on the basis of data from a previous study. For bonus points, Ss were accordingly required to give correct answers especially quickly. An extreme instance of unusually small preset values was Set 3.

Sets 6 and 7 had preset values which were slightly greater than those estimated as average on the basis of previous work with these materials in a composed-answer mode.

RESULTS

Major results consisted of printing-counter records of the time, in seconds, spent reading and answering each question. In addition, cumulative records provided a record of correct answers and correct answers which were scored as "fast," as well as a gross record of S's performance to which they could be related. In these cumulative records a millimeter step marked the beginning of each question, a half millimeter step marked when the program answer was exposed for S's viewing, another full-millimeter step, the next question, etc. Momentary displacement of the stepper pen indicated points in the #2 counter, and a special event pen, points in the #1 counter. Another special event pen marked off, redundantly, the beginning and end of each question. Throughout the study, volunteered student comments were recorded; at the end of the study, interview data were obtained.

For each S there were read-and-answer times for 10 frames in the control condition (A) and 10 in the experimental condition (B). Half the Ss were in the A condition during the first half of each lesson and half in the B condition. Since the same Ss provided 10 measures for each half of a lesson and since a total for the first half and for the second half would be based on an equal number of A- condition measures and B-condition measures, these counterbalanced features of the design could be used to advantage in correcting for the fact that one half of a lesson, for some inherent reason, took students longer than the other.

Counterbalancing could not be accomplished with odd numbers of Ss and without it, greater weight might be given one condition than another. Consequently, in such cases the data of one S were disregarded in obtaining the per cent of each S's total (A plus B) to be added to the A or B score for the half of the lesson found to require less time than the other half. In some early-lesson instances, the disregarded values were those of the S who dropped out of the study early. In other instances, the disregarded values were obtained by use of a table of random numbers.

The criterion measure obtained for each S, after the approximate correction of unequal times for lesson halves, was a difference value (A - B) based on the difference between total read-and-answer time spent on the last 10 frames in the control condition and in the treatment condition. If there were no differences, these values should be distributed around a mean of zero.

Kinds of Analyses

The effectiveness of the experimental treatment compared with the control condition on read-and-answer times was considered in relation to (a) repetitions of lessons, (b) the presence or absence of instructions concerning how S could obtain bonus points following quickly-given correct answers, (c) the number of correct answers, (d) the number of reinforcements, (e) the relative size of intervals defining fast question answering, (f) whether failures to obtain reinforcements caused the intervals to enlarge to initial size, and (g) the average read-and-answer times associated with frames on which S obtained a bonus point for speed. Also, composed-answer and keyboard response modes were compared.

Repetitions of lessons. Table 1 provides a general overview of the major findings. With the exception of the starred values, all mean differences are based on the data of 6 Ss. It is evident that the experimental treatment is typically most effective for Instructed Ss and for the first scoring for speed. The treatment was especially effective for Sets 6 and 7. That this is largely owing to the first scoring for speed on the first presentation of these lessons is further suggested by the results for Set 1, which because of several special circumstances were not included in Table 1 (cf. Table 2, Appendix). The mean for Set 1 was 16.54. When Instructed and Non-Instructed Groups were combined the means for Sets 6 and 7 were 16.66 and 16.58, respectively. (It may be recalled that Sets 2 through 5-2 were completed in the composed-answer mode once before they were completed using the keyboard, which permitted scoring for speed.)

Additional data, not shown, obtained for additional presentations of some lessons for some Ss, suggest that the ineffectiveness of the

Table 1

Mean Differences in Read-and-Answer Times Between Control and Experimental Treatments for Instructed and Non-Instructed Subjects

Time of Scoring	Instructed or Non-Instructed	Set (Lesson) 2	3	4	5-1	5-2	6	7	Average Mean Difference
First Scoring for Speed	I	1.0	21.8	4.5	3.5	-1.2	24.3	30.2	12.0
	N-I	13.8	-18.3	5.3*	12.5	-8.0*	9.0	3.0	2.5
Second Scoring for Speed	I	0.7	-4.0	-7.5	-3.7	-3.8	-9.0	2.3	-3.6
	N-I	8.8	-7.7	16.7*	-5.0	3.7*	-8.5	-14.2	-0.8
1st & 2nd Scoring for Speed	I	1.7	17.8	-3.0	-0.2	-5.0	15.3	32.5	8.5
	N-I	22.7	-26.3	22.0*	7.5	-4.3*	0.5	-11.2	1.5

Note: Positive values indicate shorter times for experimental treatment than for control treatment.

* Based on 5 Ss

treatment, in terms of difference in scores between control and treatment, persists during two or three more presentations of the lesson. Cumulative records suggest this net ineffectiveness to be more a matter of instances of unusual effectiveness cancelled by instances of unusual ineffectiveness than simple, decreased overall treatment effectiveness.

Effect of first instances of experimental treatment on criterion time measures. It was noted above that Sets 1, 6, and 7 differed from Sets 2 through 5-2 in that these three lessons were associated with the treatment-control conditions during initial presentation. Sets 2 through 5-2 were first completed in the composed-answer mode. The means for Sets 1, 6, and 7 for Instructed and Non-Instructed Groups combined were observed to be remarkably similar. Each of these means may be considered in terms of a t-test of the significance of the mean of differences between two measures of each individual (Walker and Lev, 1953, pp. 151-154). For Set 1, $t = 1.67$, $df = 10$, $p < 0.10$. For Set 6, $t = 2.53$, $df = 11$, $p < 0.025$; and, for Set 7, $t = 2.05$, $df = 11$, $p < 0.05$. (All p-values are for a one-tailed test.) On the average each frame was completed in about 1.6 or 1.7 seconds less during treatment condition than control condition, and this difference appears to be statistically reliable.

A similar consideration of Sets 2 through 5-2 inclusive suggested that the experimental treatment was not reliably different from the control condition when Instructed and Non-Instructed Groups were combined and lessons considered individually.

Statistical reliability of the difference between treatment and control conditions is evident when Sets 2 through 7 inclusive are considered from the standpoint of two kinds of overall tests. Total difference scores for each lesson were obtained which are equivalent to adding the pairs of means together separately for each lesson for Instructed and Non-Instructed Groups for the first scoring for speed (Table 1). These seven difference values were considered in a t-test for the suggestion this would have concerning the effectiveness of the experimental treatment in terms of generalizing across lessons. Accordingly, $t = 2.50$, $df = 6$, and $p < 0.025$.

A similar method was followed in considering the generality of experimental treatment effectiveness across Ss. The difference values for each lesson were collected into a total difference value for each of the twelve Ss. Accordingly, $t = 2.38$, $df = 11$, and $p < 0.025$.

These results indicate that the experimental treatment was reliably effective for most lessons for most Ss when results obtained for Instructed and Non-Instructed Ss are combined for lessons on which the treatment was first applied during the first presentation or during the second presentation. However, from Table 1, it is evident that the treatment was most effective for Instructed Ss and for the application of the experimental treatment during Ss' first encountering of the lesson.

Effectiveness of instructions. The mean values for Instructed Ss for the first scoring for speed (first row of Table 1) are all positive values with the exception of 5-2. (The mean for Set 1 for these same Ss was also positive and properly belongs in the analysis of Instructed Group results.)

The conclusions stated above of statistically reliable generality across lessons and across Ss, which were based on the results of Instructed and Non-Instructed Ss combined, are supported when only the data provided by the Instructed Group are considered. Across all lessons (Sets 1 through 7) $t = 2.90$, $df = 7$, $p < 0.025$. For Ss, of this Instructed Group, $t = 6.60$, $df = 5$, $p < 0.005$.

Conclusions concerning individual lessons (Sets 6 and 7) held for the Instructed Group, also. For Set 6, $t = 5.75$, $df = 5$, $p < 0.005$. For Set 7, $t = 3.49$, $df = 5$, $p < 0.01$. While the results of the Non-Instructed Group can be combined with these of the Instructed Group and support essentially the same conclusions, the experimental treatment would not be found to be reliably different from the control condition if only results from Non-Instructed Ss had been obtained in this study.

Despite the obvious differences between the mean-differences of 1st-row entries of Table 1, Instructed and Non-Instructed Groups were not found to differ significantly in terms of various statistical analyses (analysis of variance, a t-test based on differences between group totals for each lesson, and a t-test based on independent groups for individual lessons).

These analyses were all concerned with the statistical reliability of differences between the scores from Instructed and Non-Instructed Groups in terms of central tendency. Such differences were substantial but not statistically reliable because of the extreme variability of values within the Non-Instructed Group during this first scoring for speed.

The difference between the range of differences in read-and-answer times between control and experimental treatments for the Non-Instructed Group and the range of these values for the Instructed Group, first scoring of seven lessons considered in Table 1, was found to be statistically significant: $\underline{t} = 3.64$, $\underline{df} = 6$, $\underline{p} < 0.01$. Although the Instructed and Non-Instructed Group do not differ reliably in central tendency, the ranges of criterion values are greater for the Non-Instructed Ss to a statisically significant degree. (Substantially the same conclusion should be reached concerning differences in variability between groups with an F-test based on the ratio of within-group variances.)

Alternative corrections for inequality of read-and-answer times associated with the first and second halves of each lesson. In all of these analyses, and in others to follow, the adequacy of the correction for inequality of read-and-answer times of the two lesson halves is of basic importance. In Table 1 the correction was based on the per-cent difference between first and second halves of each lesson for read-and-answer times, which were derived from equal numbers of Ss in treatment and control conditions. Two alternative correction procedures were possible. One of these was applied only in the case of Sets 6 and 7, and the other was applied in the case of Sets 2, 3, 4, 5-1, and 5-2.

The alternative correction in the case of Sets 6 and 7 was based on the use of results from a previous study (Briggs, Campbell, and Brooks, 1964). In this previous study, Ss completed these lessons under a single experimental condition, merely responding to all blanks in program frames in a composed-answer mode. Therefore, the per cent of total read-and-answer time associated with the criterion-frame portions of the two lesson halves (last 10 frames in each half) could be compared and the per-cent difference obtained. This per cent of Ss' totals was the amount to be added to the lesson half identified on the basis of the previous results as associated with less than 50 per cent of the total read-and-answer time.

The other alternative correction procedure (which was applied to Sets 2, 3, 4, 5-1, and 5-2) was based on results obtained during the first presentation of these lessons which Ss completed in a composed-answer mode. The same procedure was followed as for Sets 6 and 7. The difference between these procedures was that the same Ss provided data for the correction and data to which the correction applied in the case of this second alternative correction. By contrast, one sample of Ss from a previous study provided data for establishing the per cent of lesson-half inequality for Sets 6 and 7 (since Ss in the present study did not complete these lessons under a single, composed-answer experimental condition) and another sample of Ss (those of the present study) were the source of the data to which the correction was applied.

Table 6 of the appendix contains the results of the present study corrected for inequality according to these procedures. Most conclusions are unaffected by the method of correction. Incidentally, other tables of this appendix give some of the details basic to results which were presented above. The first two tables are redundant for the sake of clarifying, in detail, elements of various analyses.

Number of correct answers. It seemed possible that procedures which affected S's speed of answering could be disruptive. All correct answers were considered for both presentations of lessons, first and second, in which there was a comparison of treatment and control conditions. The 4,468 correct answers occurred almost equally during treatment and control conditions: 2,232 compared with 2,236 for treatment and control respectively.

It was especially interesting that there seemed to be a greater number of correct answers the more effective the treatment was in comparison with the control (A minus B read-and-answer times): $R_s = .60$ for Sets 1 through 7 inclusive (with Set 6 missing owing to recording failure), $N = 7$. ($R_s = .96$ with Set 5-2 excluded, $N = 6$, $p < 0.01$, one tailed.)

Reinforcement, contingent upon both speed and correctness of answers, could have affected both of these properties of students' responses. The rank-order correlation coefficient of the average number of points in the #2 counter with the average number of questions correct, based on data from all *Ss* of both groups, with sets or lessons as N (and Set 6 not available owing to recording-of-correct-answers failure), was $R_s = .26$, $N = 7$, and not significant. Unusually few reinforcements occurred during Set 7, owing to the non-reset feature when *Ss* failed to obtain reinforcement. With Set 7 excluded $R_s = .93$, $N = 6$, and $p < 0.05$.

In analyses, such as this one, where the statistical significance depends upon the inclusion or exclusion of a single case, the conclusion remains in doubt, to be resolved in further research. The case built here has been that, if anything, the treatment, properly applied (when material is first presented), instead of disrupting performance is associated with an increase in the number of correct answers along with the increase in speed of answering.

It should be emphasized that under conditions in which the treatment was not effective in increasing question-answering speed, it did tend to be associated with fewer correct answers in the treatment than control condition. The *S* who dropped out of the study after the third lesson had the greatest difference in number of correct answers between treatment and control conditions and, during each lesson, gave more correct answers during control than treatment conditions.

Number of bonus points. An important correlate of criterion values, differences between control and treatment last-10-frame total read-and-answer times, was the number of points in the #2 counter obtained by *Ss*. The greatest difference between *Ss* in the Instructed and Non-Instructed Groups in terms of the number of bonus points was on Set 3, the set on which there was the greatest difference between these groups in criterion measures during first scoring for speed (Table 1). The Instructed Group compared with the Non-Instructed Group typically answered questions faster during treatment than control condition to a markedly greater extent, although this difference between groups was not statistically reliable. For each of the seven lessons considered in Table 1 the Instructed Group obtained more bonus points for quickly-given correct answers. A *t*-test applied to the differences in bonus points obtained by the Instructed compared with Non-Instructed Group yielded $t = 3.67$, $df = 6$, $p < 0.01$.

The Instructed and Non-Instructed Group differed not only in the number of bonus points, they differed significantly in the variability of the number of bonus points, with those of the Non-Instructed Group more variable. Using the range in number of such points as pairs of values for each lesson (Sets 2 through 7 inclusive), $\underline{t} = 2.05$, $\underline{df} = 6$, $\underline{p} < 0.05$.

A rank-order correlation coefficient was calculated based on the rank of criterion measures (A minus B read-and-answer times) and the rank of the number of #2-counter bonus points for each lesson, Sets 1 through 7 inclusive. Respectively, R_s-values were .74, .10, .24, .10, -.08, .10, .56, and .61. For Sets 1, 6, and 7 these R_s-values were significant, $\underline{p} < 0.01$, one-tailed. For these three lessons on which Ss were scored for speed of correct answering during the initial presentation of the lesson, the number of reinforcements was correlated with speed of answering to a statistically reliable degree. This was based on Ss of Instructed and Non-Instructed Groups combined, except for Set 1 for which essentially all Ss were instructed.

R_s-values were obtained for the Instructed and Non-Instructed Groups separately based on pairs of criterion-time and number-of-reinforcements values. For Sets 2 through 7 inclusive of the Instructed Group R_s-values were, respectively, .32, -.49, .27, -.56, .43, .43, .71. For the Non-Instructed Ss for these same lessons, respectively, R_s-values were .24, .20, .13, .46, .46, .60, and .36. It is interesting that all were positive in the case of the Non-Instructed Ss.

It seems evident that bonus-point reinforcement was as much of a determiner of speed in answering questions for the Non-Instructed as Instructed Ss. The Non-Instructed Ss merely received fewer and a more variable number of reinforcements.

Effects of preset values related to instructions. Some of the differences of criterion values among lessons is to be understood in terms of differences in preset values among lessons. For each lesson the preset values were summed and these totals compared with the time estimated for the lesson given by the program publisher. Relative to the published estimated time, the smallest preset-value total was for Set 3, and the largest was for Set 7. Excluding Sets 5-1 and 5-2, the rank-order correlation between criterion-value means for Ss of the Instructed Group and the average size of preset values was $R_s = .60$ for these five pairs of ranks, when the criterion values for Instructed Ss were either those of Table 1 or those derived by alternative corrections for inequality of lesson halves (of Table 6 of Appendix). These correlations were not statistically significant for these first-scoring data.

When the criterion-value means of Table 1 for all Ss, Instructed and Non-Instructed, are combined and are considered in relation to the relative size of preset values $R_s = .90$, and $p = 0.05$, one-tailed. The same value of R_s was obtained based on the results entailing alternative corrections for lesson-half inequality of these first-scoring data.

Sets 5-1 and 5-2 were E's arrangement of the material. Therefore estimated times were not available. The complication of multiple repetitions of material would also rule out the inclusion of Set 5-2 materials in considering rank order of criterion values. The inclusion of Sets 6 and 7 is only slightly less questionable but had dual effects on the correlation. The only imperfection in agreement between ranks based on criterion totals and those based on preset values seemed owing to the reset-non-reset difference between Sets 6 and 7. The fact that both were first presentations of material may have more to do with their greater criterion-measure mean values than the relaxed scoring-for-speed standards applied here.

There seems to be an interesting interaction between the presence or absence of instructions and the relative size of temporal interval used in scoring correct answers for speed in their effect on criterion performance. Consider Set 3 in Table 1. Although the preset values were relatively the smallest for this lesson, of any considered, the treatment, in the case of Instructed Ss, was unusually effective for a lesson involving repetition: $t = 1.90$, df,5, $0.10 > p > 0.05$.

In Table 6 of the appendix, in which the per-cent correction for inherent difference between lesson halves in read-and-answer times has been estimated on the basis of an earlier presentation to these same Ss in the composed-answer mode, the experimental treatment during Set 3 does not appear to be remarkably effective for the Instructed Group. The role of instructions is evident, however. Times appear to be reliably longer when Ss were scored for speed in the case of Ss who were not instructed in how bonus scores could be obtained: $t = 5.50$, df $= 5$, $p < 0.005$.

Accommodating individual differences. In order to shape faster question-answering performance it seemed desirable to maximize the opportunities for reinforcement. For Sets 1 through 6 inclusive each failure of S to obtain a bonus point, whether owing to his answering incorrectly or correctly but too slowly, resulted in a reset of all preprogramed comparison intervals to their largest values. Each time S obtained a bonus score for speed of correct answering all yet-to-be-applied comparison intervals became reduced. Ss had to respond correctly proportionately more quickly to obtain the next bonus point.

Set 7 was worked under the same condition of increasingly demanding scoring standards contingent upon success. However, failures to obtain a bonus score did not relax, or in any way affect, scoring standards. Instead of maximizing the opportunities for reinforcement, each reinforcement required a greater behavioral change -- increasingly faster correct answering. This procedure could be contrasted with the earlier one followed throughout the study as requiring some improvement over S's fastest correct answering rather than improvement over merely his last answering. Opportunities for reinforcement were being reduced as each reinforcement came to be that of a greater degree of performance improvement.

The Set 7 procedure appears to be superior to that of Set 6, and earlier, (Table 1) if Ss have been instructed in the relation of bonus points to their performances and if the lesson is repeated.

It is interesting that Non-Instructed Ss obtained the same number of bonus-point reinforcers on Set 7 that they did on Set 3, fewer than they received on any other lessons. Performance seems less impaired by this relative reduction in reinforcement when reinforcers are given on Set 7 according to the progressive-interval reduction procedure than it was by the initially-difficult-to-attain criterion of "fast" answering of Set 3.

Read-and-answer times at the moment of reinforcement. Table 7 gives, as cell entries, the average read-and-answer times in seconds of the frames on which Ss received #2-counter points (sum of read-and-answer times divided by the number of such frames). For each lesson in Sets 2 through 7, inclusive, the sum of these average values was less for the

Instructed Ss than for the Non-Instructed Ss: $\underline{t} = 10.30$, $\underline{df} = 6$, $\underline{p} < 0.0005$. Instructed Ss answered faster when reinforced for speed than Non-Instructed.

The relationship between average read-and-answer time for those frames of the 10 criterion frames on which S had been scored as fast, and over-all criterion performance (read-and-answer times on last 10 frames of treatment subtracted from these of control) was considered in terms of three rank-order correlation coefficients.

In the first of these correlations, the results for Ss of both Instructed and Non-Instructed Groups combined were considered. For each S the right-column average of Table 7 of Appendix was paired with the corresponding sum in the right column of Table 3 of Appendix. There was virtually no association between these values: $R_s = 0.06$, $N = 12$, clearly not significant.

In the other two correlations considered, R_s-values were obtained separately for Instructed and Non-Instructed Ss using the pairs of the values described above for these groups combined. For the Instructed Group, $R_s = .943$, $N = 6$, and $\underline{p} < 0.05$, two-tailed test. For the Non-Instructed Ss, $R_s = -.20$, $N = 6$, not significant. Why, for the Instructed Ss, is a smaller treatment-control difference associated with faster question answering at the time of reinforcement?

The nature of the relationship of reinforcement of fastest question answering being associated with less of a difference between treatment and control conditions is to be understood in terms of a third factor, rather than as a causal relation.

The third factor is a matter of individual differences in speed of question answering. Table 8 of Appendix gives the average read-and-answer times for the 10 frames coming at the end of each half of the lesson, an average time based on 20 frames of material. These averages for each S for each lesson (Sets 2 through 5-2 inclusive) are based on the first presentation of these lessons, which were completed in the composed-answer mode. Entire frames, all blanks, were responded to during this first presentation.

Pairing values in the right column of Table 7 of Appendix with respective right-column values of Table 8 of Appendix, a rank-order correlation coefficient of .89, $N = 12$, $\underline{p} < 0.01$, one-tailed was obtained for

Instructed and Non-Instructed Ss combined. This indicates that Ss who tended to answer questions quickly during the composed-answer presentation tended to answer quickly those correctly answered frames scored as fast during the experimental treatment.

When the Instructed Group only was considered in terms of pairs of read-and-answer-composed and read-and-answer scored-as-fast values, $R_s = .66$, $N = 6$, and $p < 0.05$. For the Non-Instructed Group only for such pairs of values, $R_s = 1.00$, $N = 6$, $p < 0.01$, one-tailed.

It seems that Ss' characteristic speed of question answering outside of the experimental treatment was a better predictor of performance during the experimental treatment in the case of the Non-Instructed Ss.

In the correlation noted earlier of reinforcement of fastest question answering going with less of a control-minus-treatment difference, it appears that greater differences tend to occur between conditions for slower Ss. Using mere difference of control minus treatment read-and-answer times the experimental treatment seems more effective for characteristically slower Ss. The correlation based on pairs of right-column values of Table 8 of Appendix and corresponding right-column values of Table 3 of Appendix, for Instructed Ss only, was moderate and positive, $R_s = .49$, but not statistically reliable for this value of N ($N = 6$).

The corresponding R_s-value for the Non-Instructed Ss was -.20, also based on $N = 6$ and not statistically significant.

It is interesting that characteristic performance is relatively independent of the difference values used to gauge the effectiveness of treatment, particularly in the case of the Non-Instructed Ss.

A t-test of differences between read-and-answer times of the Instructed and Non-Instructed Groups for the five lessons, Sets 2 through 5-2 inclusive as the individuals, points up a difference which is difficult to interpret: $t = 5.56$, $df = 4$, and $p < 0.01$. Whether the Instructed Group was faster than the Non-Instructed as a result of differences in instructions or for some other reason is not clear. It is likely, however, that the two groups did differ initially as a fault of the design (working with Instructed Ss and Non-Instructed Ss on separate days to avoid contact between Ss and discussion among them).

However, since the difference between control and treatment condition tended to be greater for slower Ss in the Instructed Group, if instructions made no difference, the treatment should have been especially effective for most Ss of the Non-Instructed Group.

DISCUSSION

(Major Results)

The major objective of this study was to demonstrate that faster, correct question answering could be obtained using automated shaping procedures. This objective was attained for Ss who were scored for speed, as well as correctness, to whom the basis for scoring had been explained ("Answer correctly and quickly and continue to improve.") The greatest differences between treatment and control were found when Ss were scored for speed and correctness during the first presentation of the programed material. The number of correct answers in the treatment compared with the control condition tended to increase along with Ss' speed of question answering.

There was some question concerning the statistical significance of this correlation, since this depended upon the inclusion or exclusion of one case. Assuming significance, the nature of the relationship would remain unclear. Answering correctly more often would make reinforcement for speed more likely since both conditions had to be met for reinforcement. Answering correctly could "cause" reinforcement. The reinforcing of responses according to the two properties of speed and correctness could "cause" correctness whenever it "caused" speed.

The statistically significant relationship between speed and number of reinforcements for lessons on which Ss were exposed to the treatment during the first presentation seems to suggest that reinforcement is likely to have increased correctness along with speed. There simply was too little variance in correct answers to contribute the variance in number of reinforcements necessary for this finding.

Moore and Smith (1964) reported a lower error rate on programed lessons when students received an extrinsic reward of one cent for each correct response, in addition to knowledge of results, than when there was either no knowledge of results or knowledge of results without the extrinsic reward. They obtained their findings using an early, unpublished version of the program used in the present study.

Instructions sometimes made it possible for performance to become at least momentarily independent of the number of reinforcements. Ss of the Instructed Group were considerably faster during treatment than control conditions during one lesson (Set 3) in which Ss had to answer correctly especially quickly in order for their responses to be reinforced (Table 1). An unsatisfactory explanation would be that, although there were fewer reinforcements during this lesson the responses which were reinforced, compared with those during other lessons, were given exceptionally quickly. This would have been equally true for the Non-Instructed Ss who were markedly slower, in terms of the overall criterion, although there was a significant difference between the two groups in the number of reinforcements. In part, the difference between groups may be a matter of the number of reinforcements; but, this would not explain why Instructed Ss did especially well on a lesson on which they received, relative to other lessons, unusually few reinforcements.

Before granting instructions too exclusive a role in determining performance it should be noted that these instructions had opportunity to be equally effective both during control and treatment conditions. Instructions were a background factor which helped determine the effectiveness of treatment. Instructions were not the treatment. Ayllon and Azrin (1964) report that instructions to mental patients had no enduring effect unless performance in accord with instructions was accompanied by reinforcement.

Various observations, insufficient individually perhaps, help suggest that essentially the same factors which affected performance for the Instructed Ss were equally effective in their effect on performance of the Non-Instructed Ss. For example, the correlation between the number of reinforcements and criterion performance was as great for Non-Instructed Ss as for Instructed Ss. Individually, no one of these was statistically reliable. However, all seven R_s-values were positive for these Non-Instructed Ss and of the same order of magnitude as those for the Instructed Ss. Other relevant observations will be presented later in this report (Figure 2).

The Non-Instructed Ss did not perform in a way which resulted in their coming into contact with the experimental-variable determiners of performance to the same extent that the Instructed Ss did. For example, they received fewer reinforcements, a more variable number, and their read-and-answer times on the frames associated with reinforcement were greater. Although there were statistically reliable differences, such as these between the Instructed and Non-Instructed Groups, how much is to be attributed to instructions and how much to individual differences, such as speed of question answering outside of the experimental treatment conditions, is unclear. However, while many of the differences would favor a relatively slight response to the experimental treatment for the group referred to as the Non-Instructed Ss (fewer reinforcements, reinforcement for less speed, and greater variability) the main way in which they seemed likely to differ from the Instructed Group was that they were likely to answer less quickly. If anything, the slower Ss of the Instructed Group provided the greater differences between treatment and control conditions. Much of the automated procedure was directed toward accommodating initial differences among Ss in question-answering speed. If instructions were not important to the difference between groups, the slower Ss of the so-called Non-Instructed Group should have been unusually responsive to the treatment, as indexed by mere difference in time spent in treatment and control conditions. The receiving of fewer reinforcements would tend to be compensated for by greater responsiveness to treatment, as indexed by the difference value.

The observation that slower question-answering Ss tend to be especially responsive to the experimental treatment has implications for some broader educational issues. Some of the largest individual differences in student performances would be quick to disappear if some of the important basic contingencies were to affect all students equally. Adjusting features which first fit task requirements to individual performances and then gradually change these requirements into a definition of excellent performance -- with some regression to, or stabilized phase of, less-than-excellent-performance definitions in the process after changes have been too abrupt -- are needed on a wide scale in our educational and training practices.

The technicality of many of the issues involved in test construction, in testing and placement, and, in general, in identifying those individuals ahead of time most likely to come under the control of contingencies which are inherent in some educational or vocational circumstance, have long been recognized. Considering the differences that instructions can make during an especially demanding phase of a learning task, or that a slight change in the balance between the number of reinforcements and the quality of the performance which is reinforced can make, it is likely that few educational contingencies are the same for many individual students. The technicalities involved in first insuring the control of these contingencies by fitting them to the individual performances which are most remote and then working back to a definition of performance excellence may be no more challenging than the technicalities long recognized and dealt with in test construction.

For the reader who regards instructions as a form of complex discriminative stimulus, the similarity between instructions and prompts, or cues, is likely to be evident. An implication of the difference in effect which the progressive interval-reducing feature of the Set 7 experiment had on the performance of Instructed and Non-Instructed Ss is that performance can be affected by either prompts which act before and during performance or by reinforcing events which come after instances, and before next instances, of performance. The Instructed Group had their performance affected both by a complex stimulus which set the occasion for greater speed and correctness and, presumably, by consequences which affected the probability of later instances of successful performance each time that their performance met speed-and-correctness criteria. The Non-Instructed Group had merely the reinforcing consequences to affect their performances. As bonus points came to be more difficult to obtain, the reinforcing-consequence factor began to diminish as an influence on their performance.

Observations and interpretations of Set 7 performance would seem relevant to a basic issue in learning in general and programed learning in particular in that the less behavior is prompted, the more its establishment would seem to depend on some reinforcer-type of event (knowledge of results, stimulus change related to quality of performance, higher score, money, etc.). Whether students have a chance to compare their answers with program answers would seem to be less important in the case of a strongly-prompted teaching frame than in the case of a non-prompted review test frame.

The difference between Set 3 performance and Set 7 performance of the Non-Instructed Group suggests that for a given number of unusually few reinforcers (bonus points on those two sets relative to other lessons), a procedure which provides an increasingly more demanding definition of performance excellence establishes a greater degree of performance change per reinforcement than a procedure which begins with a relatively demanding initial definition, progresses toward an increasingly demanding definition, but relaxes back to the initial definition contingent upon insufficient performance improvement. This latter procedure, of Set 3, "wasted" reinforcement on instances of performance which met only the relaxed definitions of performance excellence.

Behavior change would seem to be a function of both the number of reinforcements and the quality of performance which is reinforced. Therefore, if the number of reinforcements is to be maximized, for the sake of a more effective and efficient procedure, then it would seem preferable to decrease the size of steps in the progression toward excellence than to have fewer steps and back up all the way to the very earliest steps each time the student fails to make the improvement necessary to completing one of these larger steps. The actual size of steps would need to be established experimentally and would be affected by instructions, prompts, student capability, and other factors.

A likely refinement that would serve to increase the number of opportunities for increasingly excellent performance to be reinforced would be to decrease size of steps in the shaping process more as the number of steps completed toward performance excellence increased.

Of near-equal importance to demonstrating an effective procedure, this study was concerned with those factors which seemed to limit the effectiveness of the experimental treatment. The most likely limiting factors have been repetition of lessons, lack of instructions concerning bases for scores, scoring standards so demanding they limit the number of successes, and the failure to relax scoring standards whenever non-instructed _S_s fail to satisfy them. It also seemed likely that for some _S_s under some conditions the experimental treatment was not sufficiently gradual. Too much improvement following each success was required for a next reinforcement.

OTHER RESULTS

A combination of cumulative records, student comments, and their post-program interview helps explain why the treatment seemed ineffective when lessons were repeated and also illustrates instances of learning without awareness within the Non-Instructed Group.

Apparent Ineffectiveness of Treatment During Lesson Repetition

Ss seemed to respond much too quickly at times when lessons were repeated to be attending to the entire frame of material. One S volunteered, "You get to where you can answer without even reading!" Apparently, answers were remembered on the basis of key words or, perhaps, prompts were learned.

Under these conditions Ss could quickly answer a series of questions without following the development of a topic or otherwise making much sense out of the material. The speed gained through these shortcuts later came to be lost when S was put in the position of agreeing or disagreeing with entire statements or in the position of responding to questions for which prompts were lacking. Encountering these, essentially out of context as a result of earlier shortcuts, Ss were observed to take longer to answer questions the second time they encountered them than the first.

The interpretation was inferred from the relating of long pauses in the cumulative record, e.g., Figure 1, to frames of the program and their characteristics. Long pauses or answering times did not seem to be accompanied by errors to the extent encountered in previous studies, or as considered in a limited way in the present study. (A correlation of .52, $df = 27$, $p < .005$, between read-and-answer times and errors, for a first presentation in the composed-answer mode, for all Ss for Set 4, was found in the present study. This particular set was selected for the sake of comparison with a result obtained in a previous study (Briggs, Campbell, & Brooks, 1964) in which an r of .57 was obtained.) In repetitions of the lesson which was answered in the keyboard-answer mode, occasional substantial pauses were to be found in records provided by Ss who answered all questions correctly. Time would seem to be a more sensitive indicator of performance disruption than error.

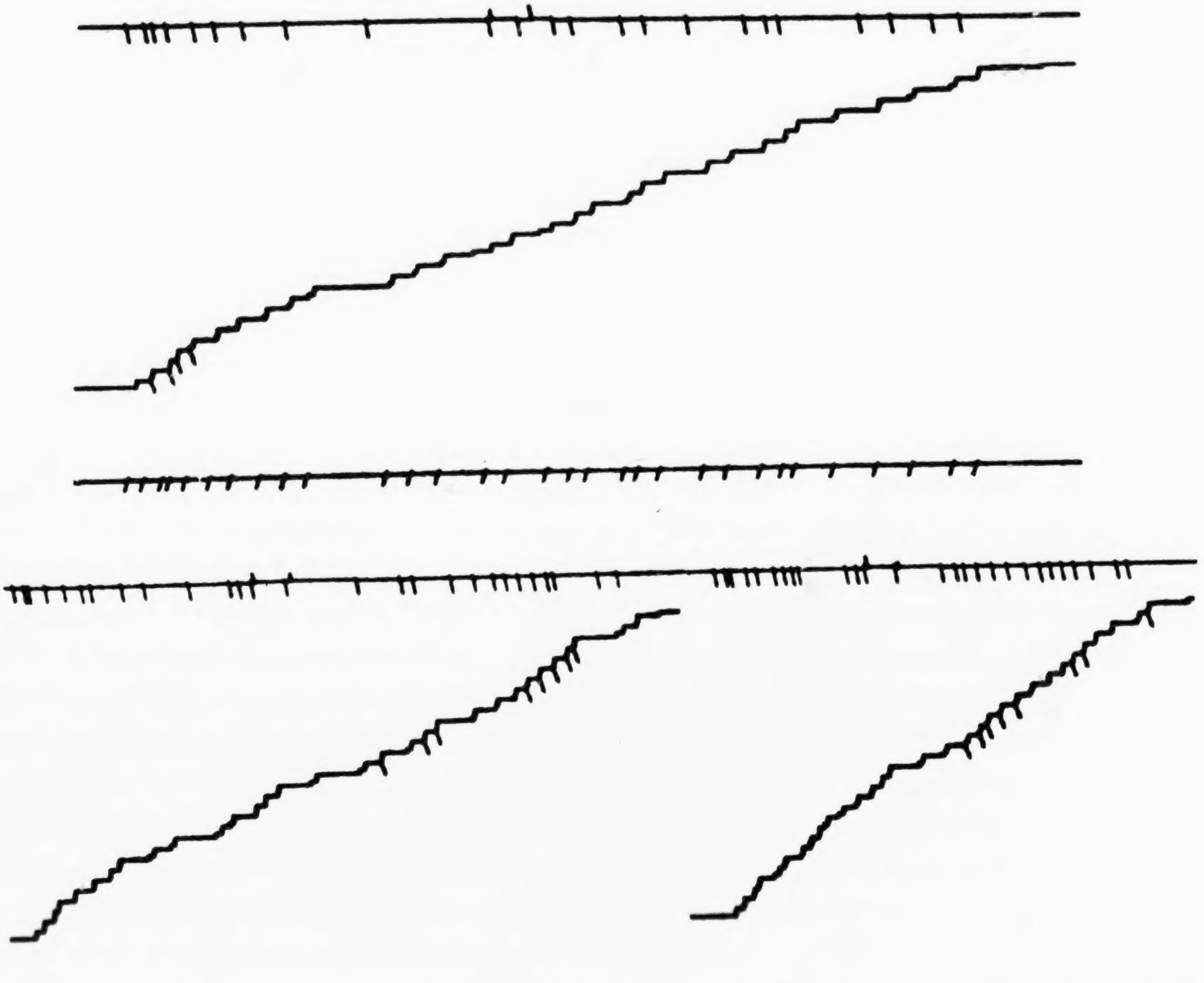


Fig. 1. The special procedure of presenting Set 7 to S #12 five times

The suggestion in the findings of this study that Ss' performance comes to be increasingly a function of prompts, or a function of limited portions of the material, with repetition of lessons would seem to have implications for the best use of programed materials. It is generally agreed that prompts are a means to the limited end of getting the desired behavior to merely occur. The greater goal is to get this behavior to occur in relation to critical properties of the material. Prompts are therefore vanished as the control of critical properties becomes established.

If the interpretation is valid that prompts are learned disproportionately as a short-cut substitute for the intended learning with lesson repetition, it would seem that a particular version of a program should be worked through only once. For review or further study, a differently-prompted version should be used -- one which had weaker prompts. Perhaps it would be most efficient, following one presentation of the program, to complete the learning by use of text material and self-tests, selected according to each student's likely difficulties.

It is probable that the conditions of the present study during the repetition of lessons were especially conducive to the "short-circuiting" of learning. Nevertheless, the relative efficiency of lesson repetition as a means of lesson mastery would seem to deserve further study. The learning which results from repetition, whether in using programs, texts, or in classroom instruction, may be especially related to increasingly narrow circumstances. This may help to explain why students can do so well on some tests and so poorly in making applications of the subject matter they seem to have learned.

Interview Data

It seemed possible that the sound of equipment action could have been important to the findings of this study. One S reported that he thought he might have been able to tell the difference between the sound of one counter (#1) and two counters operating (#1 and #2). Most Ss reported that they did not pay much attention to the counters. Some Ss of both the Instructed and Non-Instructed Groups reported being distracted or made nervous by the counters. One S reported going more by the flash of light than anything else. (But this was paired with correctness of answer rather than speed of answering during most of the study and would not explain this S's exceptionally fast question answering, treatment compared with control.) One S indicated that the sound of the equipment suggested to him how much time he would be allowed, for his correct answer to be scored as fast. (S could readily tell from the sound, which could only be partly masked, the length of the cycle. However, there was nothing to indicate how many cycles could occur before correct answers would no longer be scored as fast.)

Ss of the Instructed Group differed from the Non-Instructed Group by referring to the incentive of always trying to answer faster throughout the lesson and by referring to the competitive nature of conditions. One S commented he would much rather learn factual material from the machine than from a book, after commenting that the experiment had "stimulated a competitive instinct."

The Non-Instructed Ss reported "trying to outguess the experimenter." One S reported giving different equivalent answers and finding that for one answer he might get two points but only one point for one of the others. He reported trying to remember which had been given the first time, when lessons were repeated; and, if his first answer had only been worth one point he gave a different, but equivalent, answer.

One S reported he thought E watched from the other room and operated the #2 counter based on some movement on S's part, or that it could have been based on his speed.

Another Non-Instructed S reported that he stopped paying any attention to the #2 counter when he found it to be inconsistent. He then explained that when he answered the question the same way a second time, it wouldn't work. He thought it might be the way he pushed the buttons. "Maybe wiggling them or pushing them to one side does it."

Three of the Ss of the Non-Instructed Group referred to the possibility of the speed of answering being relevant to the #2-counter points, as one possibility. One of these reported being sometimes disturbed by the sound of equipment in that it reminded her of time passing.

In Figure 2 the positive acceleration evident in the record indicates the effectiveness of the experimental treatment. Immediately after generating this record S pointed to the #2 counter and asked, "What are the points for?"

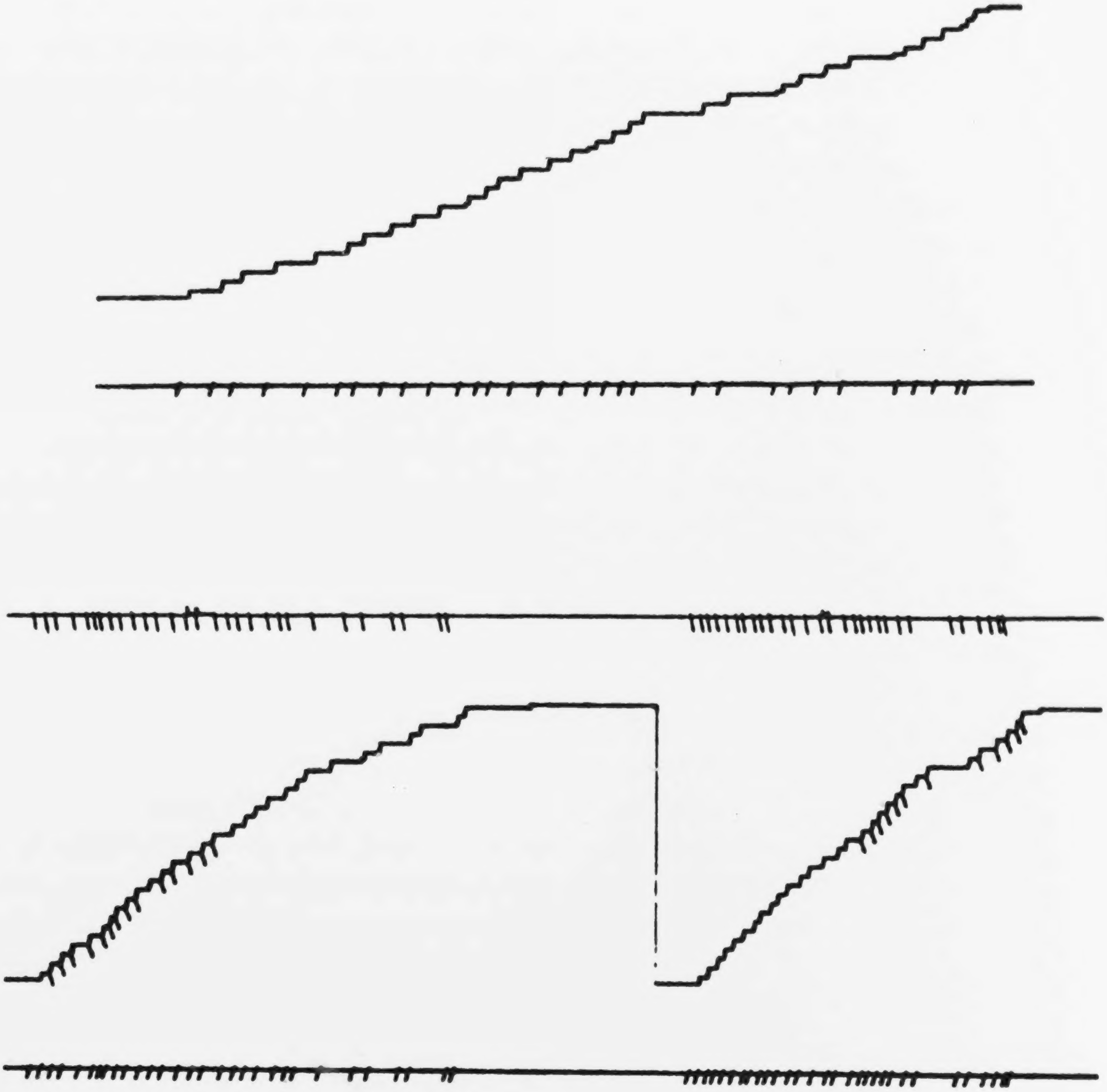


Fig. 2. Set 5-1 was presented to S #8 in composed-answer mode, followed by two presentations with machine scoring. This Non-Instructed S asked, at the end of this session, "What were the points for?"

Cumulative records, such as this one, in addition to being an objective record to compare with S's report of his performance, were useful in indicating treatment effects not otherwise evident in other criterion measures: e.g., when there was little or no difference between total times in treatment and control conditions, or when the control condition seemed superior.

Incidentally, Figures 1 and 2 represent a novel view of the learning process as "learning curves." They portray the essence of learning as an error-reducing performance-time-reducing process.

Response Modes

The present evidence from various studies concerned with response-mode comparisons, while unclear in specifying under which condition overt composed-answer responding is superior to non-overt modes in learning and retention, generally are clear in indicating that more time is required when Ss respond overtly in the composed-answer mode.

It is sometimes suggested that written responses need to be obtained in developing a program but not when the program is later used in attaining a learning objective. The objectivity of a response mode becomes a more important consideration when other aspects of performance are evaluated in addition to answer correctness, as in the present study.

It proved possible to reduce the time that would have been required for composed-answering by substituting the requirement of response to only one blank and only in terms of one letter of the word which belonged in this blank. The major drawback in using this procedure was that S was given a fixed procedure of always responding to the last blank and the first letter of his answer term. This sometimes resulted in trivial questions.

Programed materials could be prepared with several blanks which S would need to respond to covertly in arriving at this overt answer to the most significant one of these blanks, designated by some symbol. The form of his answer could be further specified in terms of which letter of his answer was to be his keyboard response, and this varied from frame to frame to insure a significant differentiation among likely answers.

Efficiency might be served by waiving even this much of an overt-response requirement at times.

In the present study Ss commented that in using the keyboard they were sometimes unsure of just how they had answered, and they sometimes preferred writing their answers for the greater opportunity it then afforded them in evaluating them. However, if the key pressed by S remained in the operated position for a moment while the program answer was exposed, this might overcome such an objection. Still, it is possible that, under some conditions, writing an answer would be of sufficient benefit to learning and retention to justify the time cost.

In the present study, the two forms of overt answering (fully composed write-in and keyboard button press) were considered in terms of the time required for each. As much as a 50 per cent savings in time for keyboard compared with write-in answering was observed (for Set 6). For Set 7 there seemed to be less than a 10 per cent difference between the two answer modes. The difference between these two lessons seems owing to the greater number of single-blank frames of Set 7. The greater efficiency of the keyboard-answer mode apparent for Set 6 seems less a difference between two forms of overt answering (keyboard compared with write-in) than it is a matter of permitting covert responding to a substantial portion of multiple-answer frames.

FURTHER RESEARCH

The keyboard response mode affords a mixture of, or a compromise between, overt and covert responding. The promptness and automaticity of machine scoring would seem to simplify the answer-evaluation mechanics of S's task. It would be interesting to study this mode further in comparison with fully-composed and covert responding in terms of the relative learning and retention occasioned by each.

A variant form of reset (accomplished, e.g., by means of a bidirectional stepping switch) adjusts preset values larger by only one unit rather than by enlarging them all the way back to initial values. Preliminary results suggest this to be superior to the reset used in the present study. Further study is needed to indicate how many instances of non-success should be the basis for relaxed scoring standards and the extent to which standards should be relaxed in each instance. Whether standards should be relaxed by either incorrect or slowly-given correct responses equally is also a question for further study.

Another future research topic concerns the level of complexity of questions used in shaping faster question answering. Since the greatest differences in time between control and treatment conditions seemed to be found for the relatively slower Ss and for first rather than second presentations of material, it is possible that relatively greater savings in time would be obtained using complex materials, having limits set only by whether Ss could answer correctly.

The possibility of evaluating and scoring for speed only an occasional correct response, perhaps a simpler procedure technically, should be considered both as a means of keeping a learning task interesting and as a means of gentle, graded, shaping of faster responding.

In the present study, preset temporal intervals were tailored to characteristics of the materials as known from previous study. This maximized the opportunities for S's correct responses to be scored as fast. Technical implementation of scoring-for-speed procedures would be greatly simplified if shaping could be as effectively accomplished by using only one size of temporal interval.

The long-term effects of the experimental treatment would also be of interest. The persistance of performances established by the treatment, as suggested in the introductory section of this report, are also of interest. However, it would be particularly interesting to consider just how radically performance could be changed through long-term differential reinforcement of faster, adaptive (correct) responding to a variety of novel circumstances.

The most effective and valuable use of programing methods and instructional devices may well be their application in developing adaptive behaviors which are not yet recognized as skills or subject matters -- where alternative instructional methods do not even exist. A generalized form of speeded decision-making is only one example.

SUMMARY

Automated procedures were considered in a series of experiments, lasting approximately one month, in which a dozen college student Ss were scored for speed of correct answering during half of each programed lesson.

Objective records were obtained both in order to determine that the automating equipment was functioning as intended, and in order to obtain some details of Ss' performance.

Ss responded to multiple presentations of the first seven sets of The Analysis of Behavior program. By pressing a lettered key of a full-alphabet keyboard, regarded as corresponding to the first letter of the last-blank answer term, Ss were able to have their answering time and answer correctness evaluated separately and the result indicated immediately in separate counters. (The fact that points were worth money if Ss remained in the study until it was completed may not have been necessary to the reinforcing effectiveness of these scores.)

Each answer scored as both correct and fast was followed by a change in scoring standards: a next score for speed required faster answering. Scoring-for-speed norms for each question were based on the results of previous research in which read-and-answer times were recorded.

The initial scoring of relatively slowly-given correct answers as fast permitted all Ss to be at least occasionally successful. By requiring increasingly faster correct answering following each success, these responses came to be differentially reinforced or shaped. During most of the study each nonsuccessful response reset the scoring standards to the initial, relaxed values. In a last experiment of the series, Ss were always required to respond correctly with increasing speed. Instances of incorrect or slowly-given correct answers had no effect on how the next answer would be scored. This last procedure had a slightly more variable effect on performance than the reset procedure.

By basing scoring intervals on times previously found to be near average for each frame of material, times which typically differed from frame to frame, the opportunity was created of being able to assess the relative speed of Ss answers to each frame. Whether such elaborate procedures are necessary or not remains for further study.

Two independent groups of six *S*s each worked under Instructed and Non-Instructed conditions concerning how bonus scores were to be obtained. Performance, statistically, was significantly less variable for the Instructed Group. Although some *S*s of both groups were responsive to the scoring-for-speed procedure (indexed by each *S*'s reduction in read-and-answer times relative to his control-condition baseline), statistically reliable effects, while present for both groups combined, were found only in the case of the Instructed Group when groups were considered individually.

For especially demanding, small, temporal-interval scoring (Set 3), Instructed *S*s were remarkably responsive to scores for speed. This suggested that the explanation of scoring, in terms of a demand for both fast correct answering and continued improvement, rendered the performances of these *S*s somewhat independent of scores for speed. The role of both instructions and congruent scoring in relation to response to scoring is clarified by recognizing that instructions, separate from the scoring-for-speed treatment, were allowed to determine the baseline comparison values of the control condition. For this reason, that instructions and scoring both contributed to faster correct answering, the effectiveness of treatment is underestimated in the analysis of results for the Instructed Group.

The statistically reliable differences between Instructed and Non-Instructed Groups may have existed at the beginning of the study, or may have resulted from characteristics which did differ. However, the nature of the difference is such as to render the Non-Instructed *S*s especially responsive to treatment (as indexed by control-treatment differences). Greatest differences between treatment and control were found among relatively slower *S*s.

Within the Non-Instructed Group, differences in performance seemed to covary with the same variables as they did in the Instructed Group, e.g., number of reinforcements.

For all *S*s, conditions which increased the speed of correct answering also seemed to increase the number of correct answers. The possibility that with more correct answers there would be more reinforcements, even if question-answering time remained the same, was examined as an explanation. The small variance in the number of correct answers is insufficient to account for the larger variation observed in the number of reinforcements based on speed of answering. The counterbalanced design in which control and treatment were present equally on both halves of lessons should also preclude all but the rarest coincidence that correct answering would just happen to be more frequent when the treatment of scoring for speed was in effect.

The role of the treatment in determining an increase in the number of correct answers warrants further study.

Half of the Ss in the Non-Instructed Group included the possibility of some kind of scoring for speed in their report of what seemed to go on during the experiment. One of these Ss seemed the least responsive to treatment, and an S who attributed scores to the way he wiggled keys when answering and made no mention of speed, was one of the most responsive. The same objective variables seemed to determine performance of both groups, but instructions seemed to increase S's exposure to these variables. The cumulative record was useful in illustrating the control of the treatment on faster performance for an S who asked immediately after completing the lesson how he had earned his score.

Cumulative records were useful as an alternative indicator of treatment effects and helped explain why the treatment seemed ineffective or disruptive when lessons were repeated. Ss appeared to speed through the program by attending to only enough of the material (key words or prompts), to be able to answer each question. Later they were at a disadvantage when the substance of the ignored material was the basis for answering certain questions. For this reason, not only should the treatment of the present study be applied during first presentations of material, in working with programed materials students probably should use materials other than the program which they have completed for further study or review.

Multiple presentation of lessons deserves consideration as an experimental technique for the study of program characteristics.

Some of the greatest absolute gains seemed to result from using the scoring-for-speed procedure with relatively slow Ss and on first presentations of lessons. Among interesting further-research possibilities would be a study of performance capability changes resulting from prolonged scoring-for-speed of Ss' responses to first-encountered circumstances to which they could just barely respond correctly.

It would seem to be as important to study ways of adjusting the contingencies of an educational or training circumstance to accommodate the initial wide range of individual differences as it would be to study ways of selecting and placing individuals in such circumstances who would be most likely to come under the control of non-adjusting contingencies. Instructions seemed to help Ss get through periods of infrequent reinforcement, with a minimum of performance decrement. A procedure which resulted in the reinforcement only of increasingly faster question answering after an initial phase of relatively relaxed scoring standards seemed superior to a procedure which was initially more demanding but relaxed to initial settings after each failure of non-instructed Ss to improve. A way of maximizing opportunities for the reinforcing of increasingly superior performance would seem to be to demand less of a behavioral change of each next improvement in performance.

APPENDIX

Effect of first instances of experimental treatment on criterion time measures. In Table 2 the corrected A-minus-B measures are given for each S for each lesson. Each number, cell entry, is the difference between the time, in seconds, spent in the control condition and treatment condition (correct answers scored for speed) based on the last 10 frames in each condition. These values were obtained after allowing for the inherent inequality of read-and-answer times on these 10 frames of the two halves of the lesson. In Table 2 the correction, or allowance, for inequality was based on data obtained during the first instance of scoring for speed of correct answering (rather than derived from data obtained either during some earlier presentation or during a prior study using these same materials). Unless there was some difference between the effect of A and B on the time spent by S reading and answering questions, each cell entry should not differ greatly from zero.

Table 2 contains all of the first-scoring-for-speed data obtained during the entire study. It includes all preliminary, or pilot, study data except for that of one S who merely helped test the equipment before the study began. The results presented in the left columns may be less definitive than those in the right side of the table. Eight lessons are considered in Table 2, since Set 5 was treated as two separate lessons.

The odd numbered Ss were told the basis for their #2-counter bonus points (Instructed Group). The even numbered Ss were not (Non-Instructed Group). The three Ss i, ii, and iii were high school students, and the details of their instruction were not recorded.

All of the results of Table 2 are for the first time S's correct answers were followed by a bonus point when given before the preset times had elapsed. In the case of Sets 1, 6, and 7, the results in Table 2 are based on a first presentation of the lesson. On Sets 2, 3, 4, 5-1, and 5-2, the first scoring for speed was after S had already completed the lesson by writing his answers in the write-in answer unit. Most of Set 5-1 was repeated again during the first-scoring for speed of 5-2.

Some data for Set 1 are missing as a result of the provisional nature of experimental procedures when the study first began. In an attempt to obtain data pertaining to a response-mode comparison, Ss in the Non-Instructed Group started on Set 1 in the composed-answer (write-in) mode. Since this took most Ss longer than the scheduling of Ss allowed, they were interrupted at whatever point in the lesson necessary to allow them time to work through the lesson twice using the keyboard. Differences in the amount of Set 1 completed before using the keyboard complicated or invalidated read-and-answer times obtained when Ss used the keyboard.

Missing data for S #13 are the result of this S's dropping out of the study, without explanation, after his third session. S #4 and S #12 both missed one session.

Each column of Table 2 represents a somewhat different condition of experimentation in terms of preset values as these relate to the estimated times given by the publisher of this program. Column 8, Set 7, differs from the others in that preset values did not adjust back to initial settings following a failure to obtain a #2-counter bonus point. Another difference among columns was mentioned above: some lessons, Sets 2, 3, 4, 5-1 and 5-2, were completed in the composed-answer mode before the presentation on which the results of Table 2 are based began. In addition, of course, the columns also differ in that results are based on different kinds of lesson content (subject matter) and on differences in programing (use of panel or exhibit material, e.g.).

Part of the greater variability of differences between treatment and control conditions of Set 1 may be owing to the randomness in length of first timing cycle, discussed above, or the differential effectiveness of the experimental treatment when applied to the first rather than second half of the first lesson. Ss who were first becoming acquainted with programed materials, with the teaching machine, with the rules for answering using the keyboard, etc., might come to be affected more by their bonus scores during the second half of this lesson than the first. An inspection of scores according to treatment versus control condition on the first half did not support this interpretation, however. A definite contribution to the greater variability of Set 1 than Set 6 or 7 is the inclusion of the three high school Ss' results in those of Set 1.

The greater variability of Set 7 cell values than of Set 6 may be owing to the progressive-interval condition during work on Set 7 (the failure to reset to initial preset values following failures to obtain points in the #2 counter). During work on Set 6, any failure of *Ss* to obtain a bonus point for speed of answering correctly, whether owing to answering incorrectly or correctly but not quickly enough, resulted in a reset of all comparison, preset, values -- enlarging them to their initial settings.

Instructed compared with Non-Instructed Ss. Table 3 contains the results presented in Table 2 in a form which makes the possibility of further analyses of these results more readily evident. Set 1 results have been deleted and the missing values for *S* #4 for Set 5-2 and for *S* #12 for Set 4 estimated by adding column and row averages and dividing by 2.

Repetitions of lessons. Some of the differences among lessons in Tables 1, 2, 3, and 6 are related to whether *S* was scored for speed on the first presentation of the lesson or only after he had first completed it by answering each entire frame in the composed-answer mode. Only when *Ss* were scored for speed of correct answering during the first presentation of lessons, in the case of Sets 1, 6, and 7, was it possible to demonstrate statistically reliable treatment effects, and then only for Sets 6 and 7. Set 5-2 is the most extreme instance of the treatment's ineffectiveness (or of its disruption). *Ss* answered questions slightly faster during the absence of the treatment (during the control condition). The lesson identified as Set 5-2 actually included most of 5-1 which was thus repeated again. Set 5-2, therefore, contained the most repetition of material of any lesson.

Table 4 presents results for a presentation of each lesson during the same session and immediately following the presentation on which the results of Tables 1, 2, and 3 are based. It is evident from Table 4 that the treatment was markedly less effective this time than previously in terms of the difference between control and treatment condition read-and-answer times. Additional data, not shown, obtained for additional repetitions for some lessons for some Ss, suggest that the ineffectiveness of the treatment, in terms of difference scores between control and treatment, persists during two or three more presentations of the lesson.

Table 5 contains results of Table 3 when added to corresponding results of Table 4. Although the treatment in the case of Instructed Ss was not reliably different from the control for most lessons when considered individually or as a group, in terms of their seven column sums ($t = 1.61$, $df = 6$, $p > 0.05$), for all lessons combined, row totals, the treatment was effective overall for all six Ss to a statistically reliable extent: $t = 3.92$, $df = 5$, $p < 0.01$. For one individual lesson, Set 7, the treatment was reliably effective for this Instructed Group: $t = 2.39$, $df = 5$, $p < 0.05$.

Allowance for unequal read-and-answer times for criterion frames of the two lesson halves. In all analyses the adequacy of the correction for inequality of read-and-answer times of the two lesson halves is of basic importance. In Tables 1, 2, 3, 4, and 5 the correction was based on the per-cent difference between first and second halves of each lesson for read-and-answer times, which were derived from equal numbers of Ss in treatment and control conditions. An alternative possible in the case of Sets 6 and 7 was to determine the per-cent difference between first and second halves for reading-and-answering times of Ss of a previous study who completed these lessons under a single experimental condition of composed-answer responding. A shortcoming of this procedure was that the data of one sample of Ss was the basis for a correction procedure for a different sample of Ss. For Sets 2, 3, 4, 5-1, and 5-2 the alternative procedure was similar. The per-cent difference between read-and-answer times on first and second halves of these lessons could be estimated from the difference between halves for these times when Ss first worked through these sets during the first presentation wherein they were completed by composed-answer responding. A shortcoming of this correction procedure is that the per-cent difference based on a first presentation of these lessons, in which they were completed in a different response mode, was applied in deriving a correction factor for second-presentation results. The cell entries of Table 6 were derived by these alternative methods of eliminating differences in read-and-answer times between totals for the last 10 frames of the two halves of lessons which had nothing to do with the effects of the experimental conditions.

Table 2

Difference Scores
(Control Condition Minus Experimental Treatment)
for First Scoring of Correct Answers for Speed
During a First or Second Half of the Lesson

S#	Sets: 1	2	3	4	5-1	5-2	6	7	Mean
1	-29	13	47	44	9	-13	19	36	15.75
3	-10	3	55	-6	36	-3	29	-6	12.25
5	50	-9	-23	-39	14	34	35	30	11.50
7	-5	2	13	-4	-25	-11	29	31	3.75
9	45	-3	13	-7	-19	2	28	60	14.87
11	35	0	26	39	6	-16	6	30	15.75
13	-8	-44	-21						-24.33
2		5	-2	7	50	18	-29	-43	0.85
4		-27	-102	51	32		-29	-9	-14.00
6		37	-7	2	-73	-61	33	-4	-10.42
8		31	-58	-6	52	12	34	41	15.14
10	42	8	1	-31	35	5	25	22	13.37
12		29	58		-21	-13	20	11	14.00
i	62								62.00
ii	-27								-27.00
iii	27								27.00
Mean	16.54	3.46	0	4.54	8.00	-3.83	16.66	16.58	

Table 3

Difference Scores
(Control Condition Minus Experimental Treatment)
for First Scoring of Correct Answers for Speed
During a First or Second Half of the Lesson
(Results Arranged for Major Comparisons)

Group								
Sets:	2	3	4	5-1	5-2	6	7	Sum
Instructed Subjects								
1	13	47	44	9	-13	19	36	155
3	3	55	-6	36	-3	29	-6	108
5	-9	-23	-39	14	34	35	30	42
7	2	13	-4	-25	-11	29	31	35
9	-3	13	-7	-19	2	28	60	74
11	0	26	39	6	-16	6	30	91
Sum:	6	131	27	21	-7	146	181	505
Non-Instructed Subjects								
2	5	-2	7	50	18	-29	-43	6
4	-27	-102	51	32	-9*	-29	-9	-93*
6	37	-7	2	-73	-61	33	-4	-73
8	31	-58	-6	52	12	34	41	106
10	8	1	-31	35	5	25	22	65
12	29	58	9*	-21	-13	20	11	93*
Sum:	83	-110	32*	75	-48*	54	18	104*
Total:	89	21	59*	96	-55*	200	199	609*

*estimated value (two estimated values affect 8 sums.)

Table 4

Difference Scores
(Control Condition Minus Experimental Treatment)
for Second Scoring of Correct Answers for Speed
(Third Presentation of Sets 2, 3, 4, 5-1, and 5-2)

Group Sets:	2	3	4	5-1	5-2	6	7	Sum
Instructed Subjects								
1	-1	-4	-32	10	14	-14	-1	-28
3	-4	2	-8	-17	1	-32	15	-43
5	-8	30	-2	1	-39	4	15	1
7	6	0	7	5	13	-8	-3	20
9	6	-22	-13	-6	-1	17	-7	-26
11	5	-30	3	-15	-11	-21	-5	-74
Sum:	4	-24	-45	-22	-23	-54	14	-150
Non-Instructed Subjects								
2	19	-16	-3	-24	-37	-26	-3	-90
4	21	23	45	-32	4*	82	-92	51*
6	6	8	8	15	35	-48	1	25
8	34	-14	14	19	23	-23	-15	38
10	-14	4	33	-8	-6	-27	16	-2
12	-13	-51	3*	0	3	-9	8	-59*
Sum:	53	-46	100*	-30	22*	-51	-85	-37*
Total:	57	-70	55*	-52	-1*	-105	-71	-187*

*estimated value

Table 5

Difference Scores
(Control Condition Minus Experimental Treatment)
for First Scoring for Speed Added to Corresponding Difference Scores
for Second Scoring for Speed

Group Sets:	2	3	4	5-1	5-2	6	7	Sum
Instructed Subjects								
1	12	43	12	19	1	5	35	127
3	-1	57	-14	19	-2	-3	9	65
5	-17	7	-41	15	-5	39	45	43
7	8	13	3	-20	2	21	28	55
9	3	-9	-20	-25	1	45	53	48
11	5	-4	42	-9	-27	-15	25	17
Sum:	10	107	-18	-1	-30	92	195	355
Non-Instructed Subjects								
2	24	-18	4	26	-19	-55	-46	-84
4	-6	-81	96	0	-5*	53	-101	-44*
6	43	1	10	-58	-26	-15	-3	-48
8	65	-72	8	71	35	11	26	144
10	-6	5	2	27	-1	-2	38	63
12	16	7	12*	-21	-10	11	19	34*
Sum:	136	-158	132*	45	-26*	3	-67	65*
Total:	146	-51	114*	44	-56*	95	128	420*

*estimated value

Table 6

Alternative Corrections for Inequality of Lesson Halves in Obtaining Difference Scores (Control Condition Minus Experimental Treatment) for First Scoring of Correct Answers for Speed During a First or Second Half of the Lesson

Group Sets:	2	3	4	5-1	5-2	6	7	Sum
Instructed Subjects								
1	24	-13	41	10	-31	40	34	105
3	-8	26	-4	35	1	17	-5	62
5	2	0	-40	16	20	50	29	77
7	-4	1	3	23	-3	22	31	73
9	8	27	-8	-17	-10	42	59	101
11	-13	-7	41	6	-2	-8	30	47
Sum:	9	34	33	73	-25	163	178	465
Non-Instructed Subjects								
2	-11	-37	10	49	35	-33	-50	-37
4	-25	-40	47	34	-10*	3	-3	6*
6	31	-40	4	-75	-45	44	-6	-87
8	15	-9	-9	-54	-9	11	44	-11
10	-4	-32	-29	33	-64	9	20	-67
12	47	-18	5*	-20	-32	39	17	38*
Sum:	53	-176	28*	-33	-125*	73	22	-158*
Total:	62	-142	61*	40	-150*	236	200	307*

*estimated value

Table 7

Average Read-and-Answer Times for Frames Answered Correctly by S and Followed by #2-counter Points During the First Scoring for Speed

Group Sets:	2	3	4	5-1	5-2	6	7	Average for 7 Sets
Instructed Subjects								
1	7.90	12.88	9.84	7.58	12.64	8.00	10.00	9.83
3	10.36	8.41	8.83	8.30	9.65	10.22	8.40	9.17
5	5.88	10.23	10.75	7.30	7.62	5.30	8.12	7.88
7	5.22	4.38	5.30	4.71	4.95	5.92	7.00	5.35
9	7.80	8.00	7.07	7.85	8.32	5.46	8.77	7.61
11	11.15	9.44	7.69	8.46	7.45	10.81	8.40	9.06
Average for 6 Instructed Subjects	8.05	8.89	8.25	7.37	8.44	7.62	8.45	
Non-Instructed Subjects								
2	17.25	14.86*	12.00	10.00	10.26	18.00	17.80	14.31
4	24.40	25.00	15.18	13.33	14.70*	12.66	15.66	17.28
6	19.50	10.66	9.84	10.44	9.13	5.53	8.83	10.56
8	11.33	16.80	12.66	9.57	13.50	15.14	14.00	13.29
10	11.91	9.63	10.16	10.07	13.89	13.27	9.87	11.26
12	12.25	15.00	11.93*	13.14	11.72	8.00	11.25	11.90
Average for 6 Non-Instructed Subjects	16.11	15.32*	11.96*	11.09	12.20*	12.10	12.90	

*estimated value

Table 8

Average Read-and-Answer Times from 20 Lesson Frames, on which Criterion Mean Differences were Later Obtained, when Completed in Composed-Answer Mode During First Presentation

Group Sets:	2	3	4	5-1	5-2	Average for 5 Sets
Instructed Subjects						
1	22.15	31.45	28.65	24.65	31.10	27.60
3	19.85	24.45	26.55	19.55	20.40	22.16
5	19.20	24.90	25.65	20.25	28.50	23.70
7	20.90	19.50	22.15	19.30	28.40	22.05
9	21.30	18.80	20.05	23.35	20.50	20.80
11	23.30	39.55	28.65	29.25	28.95	29.94
Average for 6 Instructed Subjects	21.11	26.44	25.28	22.72	26.30	
Non-Instructed Subjects						
2	27.95	40.20	41.00	29.50	36.40	35.01
4	37.85	67.55	53.75	38.85	44.30*	48.46*
6	24.05	31.00	30.10	19.35	24.90	25.88
8	27.95	34.85	38.95	24.10	31.60	31.49
10	23.25	30.30	26.95	25.30	30.85	27.33
12	28.25	36.20	31.00	23.35	32.50	30.26
Average for 6 Non-Instructed Subjects	28.21	40.01	36.95	26.74	33.42*	

*estimated value

REFERENCES

Alter, Millicent. Retention as a function of length of retention interval, intelligence, and training time. Journal of Programed Instruction, 1963, II, 7-17.

Alter, Millicent, & Silverman, R. E. The response in programed instruction. Journal of Programed Instruction, 1962, I, 55-78.

Archer, N. S., & Sanzotto, S. M. Administrative and instructional adjustments resulting from the use of programed materials. Audiovisual Instruction, November 1964.

Ayllon, T. & Azrin, N. H. Reinforcement and instructions with mental patients. Journal of the Experimental Analysis of Behavior, 1964, 7, 327-331.

Bekesy, G. V. A new audiometer. Acta. Oto - Laryngologica, 1947, 35, 411-422.

Berliner, D. C., Angell, D., & Shearer, J. W. Behaviors, measures, and instruments for performance evaluation in simulated environments. Palo Alto, California: American Institute for Research, August 1964, AIR-D81-8/64-TR.

Blough, D. S. A method for obtaining psychological thresholds from the pigeon. Journal of the Experimental Analysis of Behavior, 1958, 1, 31-43.

Blough, D. S. Scotopic spectral sensitivity in the monkey. Science, 1963, 139, 493-494.

Briggs, L. J., Campbell, V. N., & Brooks, L. O. Investigations of thinking via self-instructional programs. Palo Alto, California: American Institute for Research, June 1964, AIR-D41-6/64-FR.

Briggs, L. J., Goldbeck, R. A., Campbell, V. N., & Nichols, D. G. Experimental results regarding form of response, size of step, and individual differences in automated programs. In J. E. Coulson, (Ed.), Programmed learning and computer-based instruction. New York: John Wiley & Sons, 1962.

Briggs, L. J., & Hamilton, Nancy R. Meaningful learning and retention: practice and feedback variables. Review of Educational Research, 1964, XXXIV, 545-558.

Brooks, L. O. Response latency in programmed learning: latency related to error rate. Unpublished doctoral dissertation, University of Houston, 1961 (now available through the services of University Microfilms, Ann Arbor, Michigan).

Brooks, L. O. Student response: a guide in learning research. Paper presented at American Psychological Association, September, 1964 (American Institute for Research, Palo Alto, California, Mimeo).

Burton, B. B., & Goldbeck, R. A. The effect of response characteristics and multiple-choice alternatives on learning during programed instruction. San Mateo, California: American Institute for Research, February 1962, AIR-C28-2/62-TR4.

Clapp, Doris J., Yens, D. P., Shettel, H. H., & Mayer, Sylvia R. Development and evaluation of a self-instructional course in the operational training capability query language for System 473L, U. S. Air Force Headquarters. American Institute for Research, Pittsburgh, Pennsylvania, Interim Report No. ESD-TDR-64-662.

Coulson, J. E. Programed learning and computer-based instruction. New York: John Wiley & Sons, 1962.

Crossman, E. Unpublished doctoral dissertation, Arizona State University, 1965.

Cummings, Allana, & Goldstein, L. Effect of overt and covert responding on two learning tasks. Technical Report 620919. New York: Center for Programed Instruction, 1962 (Offset).

Edwards, A. L. Experimental design in psychological research. New York: Rinehart & Co., Inc., 1950.

Eigen, L. D., & Margulies, S. Response characteristics as a function of information level. Journal of Programed Instruction, 1963, II, 45-54.

Ferster, C. B., & Skinner, B. F. Schedules of reinforcement. New York: Appleton-Century-Crofts, 1957.

Field, G. E., & Boren, J. J. An adjusting avoidance procedure with multiple auditory and visual warning stimuli. Journal of the Experimental Analysis of Behavior, 1963, 6, 537-543.

Findley, J. D. Preference and switching under concurrent scheduling. Journal of the Experimental Analysis of Behavior, 1958, 1, 123-144.

Goldbeck, R. A. The effect of response mode and learning material difficulty on automated instruction. Santa Barbara, California: American Institute for Research, September 1960, AIR-C28-60-IR-124.

Goldbeck, R. A., & Briggs, L. J. An analysis of response mode and feedback factors in automated instruction. In W. I. Smith & J. W. Moore, (Eds.), Programmed learning. Princeton, New Jersey: Van Nostrand, 1962.

Goldbeck, R. A., & Campbell, V. N. The effects of response mode and response difficulty on programed learning. Journal of Educational Psychology, 1962, 53, 110-118.

Goldbeck, R. A., Campbell, V. N., & Llewellyn, Joan E. Further experimental evidence on response modes in automated instruction. Santa Barbara, California: American Institute for Research, December 1960, AIR-C28-60-IR-132.

Goldiamond, I. Ongoing visual monitoring: procedures for experimental analysis and control. I. Machine definition of ongoing silent and oral reading rate. Technical Report No. ESD TR61-22.

Hamilton, Nancy R. Effects of logical versus random sequencing of items in an autoinstructional program under two conditions of covert response. *Journal of Educational Psychology*, 1964, *55*, 258-266.

Hefferline, R. F., Keenan, B., & Harford, R. A. Escape and avoidance conditioning in human subjects without their observation of the response. *Science*, 1959, *130*, 1338-1339.

Herrick, R. M. Plotting and analyzing cumulative response curves in operant conditioning studies. *Journal of the Experimental Analysis of Behavior*, 1965, *8*, 59-65.

Hershberger, W. A. *Cue versus response in programed reading*. Palo Alto, California: American Institute for Research, July 1963, AIR-C28-7/63-TR6. (a)

Hershberger, W. A. *Learning via programed reading*. Palo Alto, California: American Institute for Research, July 1963, AIR-C28-7/63-TR5. (b)

Hershberger, W. A. Self-evaluational responding and typographical cueing: techniques for programing self-instructional reading materials. *Journal of Educational Psychology*, 1964, *55*, 288-296.

Hershberger, W. A., & Terry, D. F. *Complexity of typographical cueing in programed and conventional texts*. Palo Alto, California: American Institute for Research, December 1963, AIR-C28-12/63-TR7.

Hodos, W. Progressive ratio as a measure of reward strength. *Science*, 1961, *134*, 943-944.

Hodos, W., & Kalman, G. Effects of increment size and reinforcer volume on progressive ratio performance. Journal of the Experimental Analysis of Behavior, 1963, 6, 387-392.

Holland, J. G. Techniques of behavioral analysis of human observing. Science, 1957, 125, 348-350.

Holland, J. G. Human vigilance. Science, 1958, 128, 61-67.

Holland, J. G. Design and use of a teaching machine program. Paper presented at American Psychological Association, September 1960 (Harvard University, Cambridge, Massachusetts, Mimeo).

Holland, J. G., & Skinner, B. F. The analysis of behavior. New York: McGraw-Hill, 1961, 337 pp.

Holz, W. C. Decision making with free operant responses. Technical Documentary Report No. ESD-TDR-64-449.

Holz, W. C., & Robinson, J. S. A note on machines in a technology of education. Journal of Programed Instruction, 1963, II (3), 31-33.

Israel, M. L. Variably blurred prompting: I. Methodology and application to the analysis of paired-associate learning. Journal of Psychology, 1960, 50, 43-52.

Jacobs, P. I. Item difficulty and programed learning. Journal of Programed Instruction, 1963, II (2), 21-38.

Joint AERA-APA-DAVI Committee. Criteria for assessing programed instructional materials (1962 interim report). Journal of Programed Instruction, 1962, II, 57-67.

Kershner, A. M. Speed of reading in an adult population under differential conditions. Journal of Applied Psychology, 1964, 48, 25-28.

Kidd, J. S., & Micocci, A. Maintenance of vigilance in an auditory monitoring task. Journal of Applied Psychology, 1964, 48, 13-15.

Kress, G. C., Jr., & Gropper, G. L. Studies in televised instruction: individualized group instruction, Report #1 - Individual differences in learning from self-paced programmed instruction. Metropolitan Pittsburgh Educational Television Station and the American Institutes for Research, November 1964.

Krumboltz, J. D., & Weisman, R. G. The effect of overt versus covert responding to programmed instruction on immediate and delayed retention. Journal of Educational Psychology, 1962, 53, 89-92.

Lumsdaine, A. A. Some conclusions concerning student response and a science of instruction. In A. A. Lumsdaine, (Ed.) Student response in programmed instruction. Washington, D. C.: National Academy of Sciences -- National Research Council, 1961.

Lumsdaine, A. A., & Glaser, R. Teaching machines and programmed learning: a source book. Washington, D. C.: National Education Association, 1960, 536-539.

Maccoby, N., & Sheffield, F. D. Combining practice with demonstration in teaching complex sequences: summary and interpretation. In A. A. Lumsdaine, (Ed.), Student response in programmed instruction. Washington, D. C.: National Academy of Sciences -- National Research Council, 1961.

McGuire, W. J. The relative efficacy of overt and covert trainee participation with different speed of instruction. Laboratory Note TARL-LN-55-46, December 1955, Lackland Air Force Base, Personnel and Training Research Center, Training Aids Research Laboratory, Texas (Mimeo).

McNeil, J. D. Programed instruction as a research tool in reading: an annotated case. Journal of Programed Instruction, 1962, I (1), 37-42.

Moore, J. W., & Smith, W. I. Role of knowledge of results in programed instruction. Psychological Reports, 1964, 14, 407-423.

Pask, G. The teaching machine. The Overseas Engineer, February 1959, 231-232.

Rothkopf, E. Z. Criteria for the acceptance of self-instructional programs. In A. E. Traxler, (Ed.), Improving the efficiency and quality of learning. Washington, D. C.: American Council on Education, 1962, pp. 30-38.

Sidman, M. An adjusting avoidance schedule. Journal of the Experimental Analysis of Behavior, 1962, 5, 271-277.

Siegel, P. S. Concept formation: response-time considerations. Psychological Reports, 1964, 435-442.

Staats, A. W., Minke, K. A., Finley, J. R., Wolf, M., & Brooks, L. O. A reinforcer system and experimental procedure for the laboratory study of reading acquisition. Child Development, 1964, 35, 209-231.

Strollo, T. R. A generalized teaching machine decision structure with application to speed reading. Bedford, Massachusetts: Air Force Electronic Systems Division, Decision Sciences Laboratory, 1964. (Technical Document Report No. ESD-TDR-64-523).

Suppes, P. Modern learning theory and the elementary-school curriculum. American Educational Research Journal, 1964, 79-93.

Tate, M. W. Individual differences in speed of response in mental test materials of varying degrees of difficulty. Educational and Psychological Measurement, 1948, 8, 353-374.

Verhave, T. Towards an empirical calculus of reinforcement value. Journal of the Experimental Analysis of Behavior, 1963, 6, 525-536.

Walker, Helen M., & Lev, J. Statistical inference. New York: Henry Holt and Co., 1953.

Williams, Joanna P. Comparison of several response modes in a review program. Journal of Educational Psychology, 1963, 54, 253-260.

Wittrock, M. C. Response mode in the programing of kinetic molecular theory concepts. Journal of Educational Psychology, 1963, 54, 89-93.

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* American Institutes for Research in the Behavioral Sciences	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP N/A

3. REPORT TITLE

SHAPING FASTER QUESTION ANSWERING

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Technical Report

5. AUTHOR(S) *(Last name, first name, initial)*

Brooks, Lloyd O.

6. REPORT DATE June 1965	7a. TOTAL NO. OF PAGES 66	7b. NO. OF REFS 68

8a. CONTRACT OR GRANT NO. Nonr-3077(oo)	9a. ORIGINATOR'S REPORT NUMBER(S) AIR-C28-6/65-TR
b. PROJECT NO.	
c. N/A	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* Technical Report 10
d.	

10. AVAILABILITY/LIMITATION NOTICES

A limited number of copies are available through the corporate author, Station A, P. O. Box 11487, Palo Alto, California 94306

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research

13. ABSTRACT

In laboratory experiments, conducted with 12 college student Ss who completed programed lessons at a teaching machine during twice-per-week sessions, Ss were automatically scored separately for correctness and for speed of correct answering; their points were later of monetary value, to Ss who continued for the month of experimentation. Ss operated the lettered key corresponding to the first letter of the last word of their answer. Read-and-answer times, obtained in a previous study, were considered in establishing circuits for scoring the speed of each correct answer in relation to individual frames of material. Answers scored as fast were followed by circuit changes which allowed proportionately less time for a next correct response to be scored as fast. Typically, failures to obtain scores relaxed scoring standards. Scoring for speed occurred during half of each of the two (at least) lesson presentations, according to a counterbalanced design. Cumulative records indicated Ss' gross performance, correct answers, speed scores, and equipment function. Printing-counter records indicated each read-and answer time in seconds. The last 10 of these times during the experimental treatment (scoring for speed) were subtracted from the last 10 during control to obtain criterion measures. To a statistically reliable degree, Ss who were instructed in the basis for scores answered questions faster during the scored-for-speed portion of lessons, particularly characteristically slower Ss during initial presentations of lessons which began with relaxed scoring standards and reset to these following a failure. Increases in the number of correct answers tended to parallel increases in speed of answering, although overall there were 2,236 correct answers recorded in the control and 2,232 in the treatment condition.

DD FORM 1 JAN 64 1473

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Automated Teaching						
Programed Learning						
Adjusting Contingencies						
Complex Performance						
Scoring for Speed						
Response Differentiation						
Non-Subject Matter Learning						
Speeded Decision-Making						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through

____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, *or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.